

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2001, 13:03:12 ; Search time 88.53 Seconds
(without alignments)
2949.726 Million cell updates/sec

Title: US-09-508-967-1

Perfect score: 12100
Sequence: 1 MATSGSGGTQDDAKHVLD.....VNKKKEFEERYPISDIWTI 2228

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_15:*

- 1: sp.archaea:*
- 2: sp.bacteria:*
- 3: sp.fungi:*
- 4: sp.human:*
- 5: sp.invertebrate:*
- 6: sp.mammal:*
- 7: sp.mhc:*
- 8: sp.organelle:*
- 9: sp.phage:*
- 10: sp.plant:*
- 11: sp rodent:*
- 12: sp.unclassified:*
- 13: sp.vertebrate:*
- 14: sp.virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	12065	99.7	2228	5	060991 plasmodium
2	5216.5	43.1	2277	5	090065 plasmodium
3	5074	41.9	2197	5	096296 plasmodium
4	4901	40.5	3078	5	026031 plasmodium
5	4817	39.8	2182	5	026034 plasmodium
6	4809.5	39.7	2710	5	09X2B8 plasmodium
7	4711	38.9	2664	5	026033 plasmodium
8	4648.5	38.4	2209	5	090066 plasmodium
9	4493.5	37.1	2163	5	09NFB6 plasmodium
10	4341.5	35.9	2169	5	097312 plasmodium
11	4302.5	35.6	2209	5	097324 plasmodium
12	4084	33.8	2647	5	090580 plasmodium
13	4073	33.7	2042	5	025766 plasmodium
14	3989	33.0	3006	5	026032 plasmodium
15	3937	32.5	2924	5	025733 plasmodium
16	3297.5	27.3	1711	5	096108 plasmodium
17	3192	26.4	2706	5	015870 plasmodium
18	2959	24.5	1729	5	025734 plasmodium
19	2863	23.7	2212	5	094657 plasmodium

20	2668	22.0	3026	5	026030 plasmodium
21	2518	20.8	2135	5	061077 plasmodium
22	1784	14.7	438	5	09Y1N7 plasmodium
23	1771.5	14.6	455	5	09Y1N6 plasmodium
24	1758.5	14.5	1327	5	09NFB4 plasmodium
25	1681.5	13.9	3542	5	09U5M2 plasmodium
26	1629	13.5	1685	5	09U4A2 plasmodium
27	1356.5	11.2	457	5	09NC63 plasmodium
28	1319.5	10.9	431	5	096294 plasmodium
29	782	6.5	440	5	096110 plasmodium
30	777	6.4	259	5	09U7K1 plasmodium
31	728	6.0	255	5	09U7H6 plasmodium
32	724.5	6.0	247	5	09U7I7 plasmodium
33	720	6.0	480	5	09NAV6 plasmodium
34	719	5.9	254	5	09U7I8 plasmodium
35	708.5	5.9	258	5	09U7I2 plasmodium
36	702	5.8	921	5	025989 plasmodium
37	695	5.7	260	5	09U7I5 plasmodium
38	689.5	5.7	251	5	09U7H5 plasmodium
39	677.5	5.6	254	5	09U7J2 plasmodium
40	661.5	5.5	247	5	09U7I9 plasmodium
41	644.5	5.3	267	5	061076 plasmodium
42	641	5.3	246	5	09U7J6 plasmodium
43	622	5.1	298	5	061061 plasmodium
44	615.5	5.1	245	5	09U7I4 plasmodium
45	614.5	5.1	236	5	09U7H7 plasmodium

ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	2228 AA.
ID	060991			
AC	060991:			
DT	01-AUG-1998 (TREMBLrel. 07, Created)			
DI	01-AUG-1998 (TREMBLrel. 07, last sequence update)			
DT	01-JUN-2000 (TREMBLrel. 14, last annotation update)			
DE	ERYTHROCYTE MEMBRANE PROTEIN 1.			
GN	FCR3S1.2-VARI.			
OS	Plasmodium falciparum.			
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.			
OX	NCBI_TaxID=5833;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=98080592; PubMed=9419207;			
RA	Chen Q., Barragan A., Fernandez V., Sundstrom A., Schlichterle M.,			
RA	Sahlen A., Carlson J., Datta S., Wahlgren M.;			
RT	"Identification of Plasmodium falciparum erythrocyte membrane protein			
RT	1 (PFEMP1) as the resetting ligand of the malaria parasite P.			
RT	falciparum."			
RT	J. Exp. Med. 187:15-23(1998).			
DR	EMBL: AF003473; AAC05730.1; "			
RL	EMBL: AF003473; AAC05730.1; "			
SO	SEQUENCE 2228 AA; 252811 MW; 5D8C8EBFA22DC8B CRC4;			
Query Match	99.7%;	Score 12065;	DB 5;	Length 2228;
Best Local Similarity	99.8%;	Pred. No. 0;		
Matches 2223;	Conservative	0;	Mismatches 5;	Indels 0;
Gaps	0;			
QY	1	MATSGSGGTQDDAKHVLD	FGOKYHDEHGAKYVSELGSLA	IILETAFYVS 60
DB	1	MATSGSGGTQDDAKHVLD	FGOKYHDEHGAKYVSELGSLA	IILETAFYVS 60
QY	61	MOTESYTELIEANSRNPKCKDKGNDVDRFSVKQAOAYDNKKMKCSIGMTCAFFRRH 120		
DB	61	MOTESYTELIEANSRNPKCKDKGNDVDRFSVKQAOAYDNKKMKCSIGMTCAFFRRH 120		
QY	121	LCNKNFPNNNSNDSSAKAHDLLAEVCMARKYEGESIKTHYPYDSKYPISDFPMCTMLAR 180		
DB	121	LCNKNFPNNNSNDSSAKAHDLLAEVCMARKYEGESIKTHYPYDSKYPISDFPMCTMLAR 180		
QY	181	SFADIDITIRGRDLVYGNKKKKKNGKTEBKEKLEQKLKELFKIHNLVDKEAKRYND 240		

181 SFAIGDIIIRDRDLYLGRKKKONGKETEREKLEQKLEIKFIKIDNKLKDEAQRNGD 240
241 EDPNFKYLRDMWMTANREVTYVGMATCSKELDNSSYFRATCNDTGGSPOTHNKCDDK 300
241 EDPNFKYLRDMWMTANREVTYVGMATCSKELDNSSYFRATCNDTGGSPOTHNKCDDK 300
301 GANAGKPRAGDGDVYTYVTFDYVPOYLRFWEEMAEDEFCRRKKKKLEMLEKQCRGKDS 360
301 GANAGKPRAGDGDVYTYVTFDYVPOYLRFWEEMAEDEFCRRKKKKLEMLEKQCRGKDS 360
301 GANAGKPRAGDGDVYTYVTFDYVPOYLRFWEEMAEDEFCRRKKKKLEMLEKQCRGKDS 360
361 EYRCSNRNGYCEOTISKKGKVRMGKCTDCFPACGSTEYNNIDNOROKPDKOKTYKIS 420
361 EYRCSNRNGYCEOTISKKGKVRMGKCTDCFPACGSTEYNNIDNOROKPDKOKTYKIS 420
421 DGGRRKRAVAGTTRYESYKSEYKLEKNDGTYVDAFLGLNNEKACKDITDGGKINFK 480
421 DGGRRKRAVAGTTRYESYKSEYKLEKNDGTYVDAFLGLNNEKACKDITDGGKINFK 480
481 EVNSGGVYVGGSGTSGASGTNDENKGTFRSEYQPCPCGYOHKGNOMERTKYK 540
481 EVNSGGVYVGGSGTSGASGTNDENKGTFRSEYQPCPCGYOHKGNOMERTKYK 540
541 MRMSKLYPRINGKVLKSLKVVYKDMILKKNKEFCLTONSSDGSYGVYTTGASG 600
541 MRMSKLYPRINGKVLKSLKVVYKDMILKKNKEFCLTONSSDGSYGVYTTGASG 600
601 SEKKELDEMKCYHNEVOKVNOGEVEDEDDDELKAGGLCILPRPKKNKEVSEAKSGNN 660
601 SEKKELDEMKCYHNEVOKVNOGEVEDEDDDELKAGGLCILPRPKKNKEVSEAKSGNN 660
661 HADIOKTFHDFEYVVAHMLKDSIHMWTRKLSISDGKTKCRNGKCKDCEKWNQ 720
661 HADIOKTFHDFEYVVAHMLKDSIHMWTRKLSISDGKTKCRNGKCKDCEKWNQ 720
721 KETEMKPKDHFKEQEGYPTTLELILKLOFLKEDTEENTENSLDAEAELEKILQ 780
721 KETEMKPKDHFKEQEGYPTTLELILKLOFLKEDTEENTENSLDAEAELEKILQ 780
781 KILLENENNLAVYVAGTEOKTLMDKLNLHNLNATKCDPLPEEDKSRRSADPSDI 840
781 KILLENENNLAVYVAGTEOKTLMDKLNLHNLNATKCDPLPEEDKSRRSADPSDI 840
841 FIPPEEKEDDENDEDEYRDEDETAKEETEGSANDTTLSDVCPYVGRVLTAKDNESL 900
841 FIPPEEKEDDENDEDEYRDEDETAKEETEGSANDTTLSDVCPYVGRVLTAKDNESL 900
901 QDASGLKYGNNRGLGRCVTPSGEPTSSDKNGAICYPPRRRLYIKIYDMATKTESP 960
901 QDASGLKYGNNRGLGRCVTPSGEPTSSDKNGAICYPPRRRLYIKIYDMATKTESP 960
961 QASGEASSTGSTTPPDSKALKAFAVESAIETFFLMHRYKEBKRAVAOEGHGLPR 1020
961 QASGEASSTGSTTPPDSKALKAFAVESAIETFFLMHRYKEBKRAVAOEGHGLPR 1020
1021 VEESPEYDEPDKLEKGIIPDGLRQMEYTLGDYRDLIFSSGNDTTSYKSDTPSSNDNL 1080
1021 VEESPEYDEPDKLEKGIIPDGLRQMEYTLGDYRDLIFSSGNDTTSYKSDTPSSNDNL 1080
1081 KNYILLASGSTEOREKKNKKKKEIKNFRKCTERSAPVLVSHPOTWNNCKYIWHGAVC 1140
1081 KNYILLASGSTEOREKKNKKKKEIKNFRKCTERSAPVLVSHPOTWNNCKYIWHGAVC 1140
1141 ALTSKDKIAKVEKPPKIEPENLMDANKPKPOYQYNNVNLDESGTSPRTOTQA 1200
1141 ALTSKDKIAKVEKPPKIEPENLMDANKPKPOYQYNNVNLDESGTSPRTOTQA 1200
1201 SSDNPTTLTHFVAPTYFRWFEEMGESFCRERKKRLKQIVDKVENGADVGRCSGDEA 1260
1201 SSDNPTTLTHFVAPTYFRWFEEMGESFCRERKKRLKQIVDKVENGADVGRCSGDEA 1260
1261 CDSISTHDYSTVPFNCBCKHCSSTYRKWTERKKIEFHKSNAVGOQKTDATRNNGTTF 1320
1261 CDSISTHDYSTVPFNCBCKHCSSTYRKWTERKKIEFHKSNAVGOQKTDATRNNGTTF 1320

1261 CDSISTHDYSTVPFNCBCKHCSSTYRKWTERKKIEFHKSNAVGOQKTDATRNNGTTF 1320
1321 DKEPCKLETWPDAAKFLERLKNGPCNKEYGDDIDFEKDSOTFOHTEYCGCPKPKT 1380
1321 DKEPCKLETWPDAAKFLERLKNGPCNKEYGDDIDFEKDSOTFOHTEYCGCPKPKT 1380
1381 NCQNGNCVSGJLNGCNDGKSIDAKELAKMRSSTDDVVRVSDNDTTFEGDLDKACQH 1440
1381 NCQNGNCVSGJLNGCNDGKSIDAKELAKMRSSTDDVVRVSDNDTTFEGDLDKACQH 1440
1441 ANIFKGIKDVWKGYCVGVICBOTNINERTDQKEYIQIATLKRVRWENLEBYNKTND 1500
1441 ANIFKGIKDVWKGYCVGVICBOTNINERTDQKEYIQIATLKRVRWENLEBYNKTND 1500
1501 KISHCIRKGEKSCINGCEKSKCEKIEKKIEMERIKRFDQYENKQOPDYNKSI 1560
1501 KISHCIRKGEKSCINGCEKSKCEKIEKKIEMERIKRFDQYENKQOPDYNKSI 1560
1561 LEBELIPKIAVVDODNVYKLCVFENSKCTLISNTONKENDADICMLKRLGVAKNCPG 1620
1561 LEBELIPKIAVVDODNVYKLCVFENSKCTLISNTONKENDADICMLKRLGVAKNCPG 1620
1621 KPSEKOSDCKEPPPLPDEEONPEENTLEPPKCPPTTOPPEEKGETCGNKEBKDEK 1680
1621 KPSEKOSDCKEPPPLPDEEONPEENTLEPPKCPPTTOPPEEKGETCGNKEBKDEK 1680
1681 KESEEPKKEESGPAABEPAPTAASEETETNFPBPBGTPAAPSTAPPTPDPPLRP 1740
1681 KESEEPKKEESGPAABEPAPTAASEETETNFPBPBGTPAAPSTAPPTPDPPLRP 1740
1741 QADEPFDSTLIQTTIPFGVALALGSIAELFLKKTAKSVGNLFOILOPKSDYDIPPLKS 1800
1741 QADEPFDSTLIQTTIPFGVALALGSIAELFLKKTAKSVGNLFOILOPKSDYDIPPLKS 1800
1801 SNRYIPYSDRYKKTYYIYMGSDDEKYYAFMSPTDVTTSSESEYEELDINDIYVPSPK 1860
1801 SNRYIPYSDRYKKTYYIYMGSDDEKYYAFMSPTDVTTSSESEYEELDINDIYVPSPK 1860
1861 YKTLIEVYLEPSSGNNTTASGKNTPSDTRNDIONDGIPSSKITDDEMNOLKKEFTSNMION 1920
1861 YKTLIEVYLEPSSGNNTTASGKNTPSDTRNDIONDGIPSSKITDDEMNOLKKEFTSNMION 1920
1921 QPNVNDYTSGNSSTNTNTTTSRHNVDNNTNTTMSHDNNEENLLPSIHDGNLYSGEE 1980
1921 QPNVNDYTSGNSSTNTNTTTSRHNVDNNTNTTMSHDNNEENLLPSIHDGNLYSGEE 1980
1981 YSYNVNMYNSMNDIPINDDNNVYSGIDLINDLSGKRPIDYDEVLARKENELFGTEWTK 2040
1981 YSYNVNMYNSMNDIPINDDNNVYSGIDLINDLSGKRPIDYDEVLARKENELFGTEWTK 2040
2041 RSTONVAKTTPNSPIHNOLELFHKWLDHRHDMCEKMKNEEDILNKLKEENKKEINNSG 2100
2041 RSTONVAKTTPNSPIHNOLELFHKWLDHRHDMCEKMKNEEDILNKLKEENKKEINNSG 2100
2101 KTYNSDNKPSHNHVLNLDVSIQIDMDNEKTKNEITNMDTNDKSTMDTILDDLEKYNDPY 2160
2101 KTYNSDNKPSHNHVLNLDVSIQIDMDNEKTKNEITNMDTNDKSTMDTILDDLEKYNDPY 2160
2161 YYDFEEDDIYHDVDEKSSMDIYVDHNNVTSNMMVYPTFMHTEMNVNKKKEIFEERY 2220
2161 YYDFEEDDIYHDVDEKSSMDIYVDHNNVTSNMMVYPTFMHTEMNVNKKKEIFEERY 2220
2221 PISDIWNI 2228
2221 PISDIWNI 2228

RESULT 2
09U0G5 PRELIMINARY; PRT; 2277 AA.
ID 09U0G5
AC 09U0G5;
DT 01-MAY-2000 (TIREMBLrel. 13, Created)
DT 01-MAY-2000 (TIREMBLrel. 13, Last sequence update)

DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE VAR.
GN MAL4P2.58.
OS *Plasmodium falciparum*.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RA Devlin K., Bowman S., Churcher C., Harris B., Harris D.,
RA Quail M., Barrett B.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL, AL035475; CAB62899.1; -;
SQ SEQUENCE 2277 AA; 256122 MW; EA2620CA69DDEF93 CRC64;

Query Match	43.1%;	Score 5216.5;	DB 5;	Length 2277;
Best Local Similarity	48.4%;	Pred. No. 2.3e-276;		
Matches 1160; Conservative	272;	Mismatches 643;	Indels 323;	Gaps 78;

```

0Y 70 L1EANSKRNPPCKKDGKNDVDRFSYKBDAGDGNKKMSGCM-----TAPRRRLHLCN 123
Db 74 L1GDSGERHPGNTTGGKEDVDRFSYKQDAEYDNKKMKCSYSGNCTDYGACAPYRRFLCN 133
0Y 124 KMFPMNNSDSSKAKHDILAEVCMAAKKEGESITKHPKYDSKPGSDPFCMCLARSFA 183
Db 134 KMEKMKGR--ISTYKHDLLDLVCAANTEAOSLIRYHDKHOLNREGO--ICTYLARSP 189
0Y 184 DIGDIIIRGRDLYLGNKKKKKKKKGKETERELEBQKLKEIFKKIHNLMD--KEAQRYNGE 241
Db 190 DIGDIIIRGRDLYLGNKKKKKK--ETERQDLBSLKLKIFGDIYNELTGRNGVDRKHQDDN 247
0Y 242 DPNFKLEMDWTANRELYWGAAMTCSKELDNSSIFRATCNDT--GQGP-SQTHKKCRDKD 299
Db 248 GGNFYOLREDDWMTANRATVWAKITCKADTGN-AYFRPTCSDSQKGSFSAOANDCRCKK 306
0Y 300 KGAANGKPRAGDGVITYPTPEYEDVPOYLRFNEEMADDFCKKKKLEJNEKCRGDKS 359
Db 307 KKKNTDQ-----VPTTFDYVPOYLRFNEEMADDFCKKKKIYGIYKTCREKYS 366
0Y 360 -DEYRYSRNGYDCBOTISRRKGVAMGKGCJDCFFACSGYEMWIDNORQOPDKOK--KYTK 417
Db 357 GNEPRYCSNGYDCKTKTRAIKGRYMGQCSLCYLACNPYDWTNNKQBDPDKOKYDK 416
0Y 418 EI-----SDGGKRRKRAVGSTTYEGYEKSYEKLANDGYDAFLGLLNNEKACKDI 471
Db 417 EIKYIKKNGASGIRROK--GTTTKYEGYEKKFYDKLEKNNGYTGVEFLGLLNNEKACKEV 474
0Y 472 TPGGKINKEVY--SGGGV-----GGSGSGTSGASTNDEBNKGTYYRSYCO 517
Db 475 KGGGTDTRQVNVSTSGTAIVASASGASTSGGSGAASGSDTSGTNAASOGTYRSBYCO 554
0Y 518 PCPPDGVYHKGGMQWERKTYKKKRWMSKLY--KPIKGWILLKSLKVVYKDMMLIKKN 574
Db 535 PCPHGVGVRKANNGNVYKSSDBECKNITLYPRKPRREGSTKIELIKSSEGETE---IKEL 591
0Y 575 KEFCULONSSDGSVGSVYTTGASGENSEKELVDEMCKYKNEVOKVNOGVEEBDD-- 632
Db 592 EOPCOTQATG-----VANSGSGTSGSKLYEDMKY--NDEVEKGDOG--VDDDDL 643
0Y 633 ---ELKGAAGLICILPNPKKNEVSAKSONNHADIOCTFHDFFYUVAHHLKLSIHNRK 669
Db 644 YDLVNASGGLCILT--OKRNEENGKK-----OKTYNDFEFWVAHHLKLSIHNRK 693
0Y 690 RLKSCISDQTKYKNGCNKRCDCFEKVVWOKETEMPRIKDHFOTGOSIGEGYV--FTW 748
Db 694 KIKGCLAKNGKAIKCTDKKGCQCKPERVNEVOKREEDWNIKENHREKODITPGLLPDALL 753
0Y 749 LILKLOFLKEDTEENTENSLDAEAEELKHLQKILKLENNENLAVVNAAGTEOKTIDLKL 808

```

Db	754	GVLEGGVLLTSIKAYG--DAKETHEIKOL-----LDETAVAGGVVYGXKDNWTTIDKLL	805
Qy	809	NHELNDATKRCDCPLPEEDKSRGSAPSPDIFLRPEEKEDDEDND 3DEVRD-----	863
Db	806	EOELDKAKCKKCEQRKPGEEGGAARLPGVDTTVDANEDDLDDDD 3DEDDDGSGSD	865
Qy	864	-----DEBATEETTESATTTTTSLDVCPIVKGVLTKDNE5LQDA*SLKYG-GNNS	913
Db	866	VGSGSVGEVEEBTAKAEATE-----ETTRPLDYCNVTKALEGD---LGEA*ROKEYEGREK	918
Qy	914	RLGRCVTV-----PGEPT-753DKNGAICVP	940
Db	919	FPNMKICIGDMWTRGSESESAGPSRSKRHTTESDSAYATGSSGEATGKS3DKGCAICVP	978
Qy	941	RRRLRYI---KKIYDMATKTESPOASGSEASTSGSTPDPDSKALLKAFVESA4ETFE	997
Db	979	RRRLRYITLGGFRLLDGT-----AVSEATOACTPSQSPKDALLLMAFVETAA4ETFE	1033
Qy	998	LWHRKREKKKVAOE--GAGHGLPRVE---EGSPEDYDEPKLKE-KRIPDGLRQMFYTL	1051
Db	1032	LWHRKROKCKKPKYNEVGGAAGVLQITIGTLENSEQNPQKKLQESIEIPDDLRLQMFYTL	1091
Qy	1052	GDRYLIFSGSNDTTSVKDTPSSNDNLKNIVLLASGSTEOREKMKNTKEIKFRKCS	1111
Db	1092	GDRYLIVRGVAD-----DKNGNNIILMASGN---KEDMERKKKIOQEDIKI	1136
Qy	1112	TRSRAPNLVSRPO-----TWENNKGXIIWGMVVALSKKIKAGVEKKRPKEJENE	1166
Db	1137	LEKSSSEASGAQKNGSISREKMMWDKAPSTWDMGVALYNTDTASGT--APTQIOEYK	1199
Qy	1164	N-LWDEANKKKRP-----POYQYINVKL-DENSGT--SP-----FTTQOASSDNT	1205
Db	1199	NALLDGEKKRKQNGTNGKQDTYGGVVLBEDNSQTALSPNAPASNASTTQOSSISEWT	1255
Qy	1206	PTTLTH----FVKRPTYFRNFEEMGESFCERKKRLKQIKVDCKY-EG-G-DVGRCSGD	1257
Db	1255	PTTLNPKLKDQFVLPTFYRLFEEMGONFCERKKRLDQYIRECKYDETGPDRGKKCSGY	1314
Qy	1258	GPACGSISTHDSYTPSNCGCGCKHSSYRKTIERKKIEFHOSNANAYOQKTDATARRN-	1314
Db	1315	GEDCDNLSSKRTDPLPSLECGGCRHCSFKKWKIITKDDYEKOQKAYIEQRTNYNNKK	1377
Qy	1317	---GNTPFRECKTLET--WPAARFELRLKNGPKCTNKE-----YGIDDIDFEKDSKT	1361
Db	1375	VESNNHDKCECTNLETYYTAAANFLQRLADGPKNNSEEDQVNGY11FD-DISMD-KT	1433
Qy	1366	FOHTEYCGPRCEFKTNC-QNGNCVSGLNGCDDGKSIDAKETAKMRS;TTDYVMKSDN	1422
Db	1433	FGEHNYCDPSCFETVNCNRNDHCNS--NGNCKDNKITAEKIGANGVD;TV-LDMREVIDD	1488
Qy	1425	DNTNTEGGDLKDACOHANIEFGIRKDDVMKCGYGVNDICE-----QIILINETDCKEYI	1477
Db	1490	SATGGRGGLLEACGSANIFGRIKKEGMCCKGYGVWVCKPKGBNRTPIRGENDNKHII	1544
Qy	1479	QATRLCFMWEVFLDYKNIDKISHCIKKEGSKCJINGEKSKSCLEIWIIEKIAEMEN	1534
Db	1550	TTRALVTWYQNFLEDYKIKIKHKSISHTCKTDGOSTCONKQCNCKCVGMIRIKLQOEMBE	1601
Qy	1539	IKKRENDQYENKDPDYNKSIILEELIPKIAV--NDQDNVIRLCVFINSKGCTLLISMT	1591
Db	1610	IKKRFLNQYKMDSDIEYVPRSVLFTPLVQIGAANANNVDKLLIKLSBF.KSCGCSAKTNS	1661
Qy	1596	QNNKENDALDCKMLKTLGYKANKCGKSGSEKQSDCKRPPRLPDEBD-----QNPENLTL	1644
Db	1670	ENNKEDALDCKMLKTLGAKKACHDOHSNDQEKCDPPEPLDEBDLL;EEDQNPANMR-	1726
Qy	1650	BPPKPCP---PTTOPPEKGETGCKNKEEKDEKKESESEPAKEESGPAAEPPAPAESE	1706
Db	1729	--PGCFQNDTTBOOEENICTFAEYVKKEEKEKEOEDEEBDEKVP-----	1777
Qy	1707	ETETNPPPPGTGAAPPSTPAPPTPTPLPLRQADEPDSITLQI-TIPGVALATGS	1766

```
Db 1775 -----PRAPAPASKEKKEKPSQPPRRRTLELDNPHVQATLVSTLAMSVCIGFAA 1826
Qy 1766 IAPFLKKKTKASVGNLFQIOLQIPKSDYDIPFLKSSNRYIVYSRKYGKIYVWEGSD 1825
Db 1827 FTYFLKKKTKASVGNLFQIOLQIPKSDYDIPFLKSSNRYIVPASDRYKGIYVWEGSDS 1886
Qy 1826 -EDKYAFMSDDTDTYTSSESEYEELDINDIYVPGSPKYTLIEVLEPGNNMTTASGKMT 1884
Db 1887 GDEKYAFMSDDTDTYTSSESEYEELDINDIYVPGSPKYTLIEVLEPGNNMTPT----- 1940
Qy 1885 SDTRNDIONDCIPSSKITDNEWMOLKEFISMLONOPNDVPNDYTSNGSSNTNTNTTTS 1944
Db 1941 SDIPSPNSDTP-PPITDDDMNOLKDPISNMLQNTQNTPEP-----IL 1984
Qy 1945 RHVNDNNTMTMSRDNMEENLLPSIHGNTLSGSEYSYN-----NMVN----- 1989
Db 1985 HDVNDNNTHTPSRHNMQKPFILSHDRNLFSGEYIVDFMNSGNNINISDSTNSKDS 2044
Qy 1990 --SNMDIPINBDNNVSGIDLINDLSGKPIDIYDEVLKREKENLEFGENTTKRTSTON- 2046
Db 2045 LTSNNHSPYNDKNDLSGIDLINDALSNGH-IDIYDEMLKREKENLEFGQHHPKNTISNR 2103
Qy 2047 -VAKTTSNDPILHNOLEFHAKLDRHRDCEKWKAKEDILKLEBKENKENTINSKITYNS 2105
Db 2104 VVTQTSDDPITNOINLFHKMLDRHRDCEKWKAKENHERLPKLKEIM--ENETHSGDI--N 2159
Qy 2106 DNKPSHNHVLTDVSIQIDMNPKTKEITMDTNOOKSTMDTILDOLEKYNDDPYDYF 2165
Db 2160 SGITSGNNHVLTDVSIQIDMNPKTKEITMDTNPDSMTDITLDOLEKYNDDPYDYF 2219
Qy 2166 EDDITVHDVYKSSMDIYVDDHNNVTSNMDVPTKMHIEKNIVNNKKEIFEEXEYPIIS 2223
Db 2220 KHD-IYDVNDKASEDHINMDHNMKNNDNSDVPITVQIEKNVYINN-CELLONEPEPIS 2275

RESULT 3
096296 PRELIMINARY; PRT: 2197 AA.
AC 096296:
DT 01-May-1999 (Tremblrel, 10, Created)
DT 01-May-1999 (Tremblrel, 10, Last sequence update)
DT 01-May-1999 (Tremblrel, 10, Last annotation update)
DE PREMPL.
GN PFBI055C.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99021743; PubMed=9804551;
RA Gardner M.J., Tettein H., Carucci D.J., Cummings L.M., Aravind L.,
RA Koonin E.V., Shalton S., Mason T., Yu K., Fujii C., Pederson J.,
RA Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Pertea M.,
RA Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,
RA Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.;
RT "Chromosome 2 sequence of the human malaria parasite Plasmodium
RT falciparum."
RL Science 282:1126-1132(1998).
DR EMBL: AF001434; AAC71996.1;
SQ SEQUENCE 2197 AA; 249668 MM; D2A92DB54535C143 CRC64;
```

```
Qy 114 APERHLCKNKNFPMNSNDSSKAKHDLAEVMAAKYEGSEIKTHPKYDSSKGYSDPF 173
Db 123 APERHLCKSH---NLESDITTSMTKLLLEVCMMAKEGNSIDTHYQHORTEDSISQ 179
Qy 174 MCTMLARSFADIGDILIRGRDYLGNKKKKQNGKETEREKLEOKLEIFKJHDLKKE- 232
Db 180 ICTMLARSFADIGDILIRGRDYLGNKKKKQNGKETEREKLEOKLEIFKJHDLKKEG 234
Qy 223 AOKRYNGDEDEPNFKLRDWTANREIYWGAMTSCKELDNSYRATYCNLTGQGPSQTHN 292
Db 225 AETRY-GSDITNTYQLRBDMWYANRATYWEAITC--DVHGSDYPROTCGDGETTATRVKD 291
Qy 293 KCRDCKDGANAKPKADGDVTITPTFEDVYPOYLRFMEEMADFCRKKKLELEQ 352
Db 292 KCRCKDEGKKRG-----SNADVPYTFYVPOYLRFMEEMADFCRKKKLELEQ 345
Qy 353 CRGKDKDEYR---YCSHNGYDCEOTISRKGVKRGKCTCFRACSGSYENMIDNORKOF 409
Db 346 CR-----DYQNLYCSGNGYDCTRTIKYKGLVIGEHCTCSYWCRLYESMIDNOKLEF 399
Qy 410 DKOK-KTYKEISDG-----GKKRAVG--GTTKEGYEKSYEKLKNDGCTV 455
Db 400 LKOKQYETEISNGSGSGGVKGRNKKRGAGETATNYDYEKFKELKESYGV 459
Qy 456 DAFGLLNNEKACKDITD--GGKINPKEVNSGGVYVGGSGGTSGASGTNDENKGTFRSE 514
Db 460 DDFGLLNNEVCKKIDEREKIDFTK-----PADKNSNEGTFTISE 502
Qy 515 YCQPCPDGVQHGKGNQMERKTKYKMKRSKLYPRINGKVLKLSLKVNDMMILKKW 574
Db 503 YCKPCPDGVKRR--DNQKDKYDGCCTR--GKLYEPASAGQPTIKILSGEKOKEIETKL 560
Qy 575 KEPLTONSSDGSVGYTTC-----ASGNSKEKELDEMKCYKHNVEQKY-----NVQ 624
Db 561 KAFK---DQINGDITNSVARGGADSGSKSNSKELYEWKCY--NEQVKYKDKNDEE 615
Qy 625 GEVEEDELKGAGLICLTPPKKKNK EYSEAKSONNHADIOKTFPHFYYVVAHMLKDSI 684
Db 616 DEDEEDVDKVKAGGLCTLEN---KHNESRNNSNEPQPKTHDPFYFVIGFNLDSM 672
Qy 685 HWRTKRLKSCISDGTKMCRNGKNCCKDCFEKWKVQKETEKKPIKHFKTQE----- 736
Db 673 YWRGK--VNSCINNPKRKKRCKEKCDCGCFKEWIGKKKEEENIKKHFKTQEAERKNREN 731
Qy 737 -----GIRPGYFTTELILKIQFLKEDTEBMTENSLOAELELKHOKIL-----K 784
Db 732 SGIDMFSGLMDS-ADVYLELLELEQLFQDITKD-----YGVKELKIKELDEERK 784
Qy 785 LENENNLAVNAGTEOKTLMKDLNLHNLNDATKCDPLPEEDSRGSRADSPDIFP- 843
Db 785 KQAEAVVYVVAADQOKTTIDKLLOHEDDANNL-----KTHKEKEEPQPKP-----PG 835
Qy 844 -----RPEEK-----EDDENEDDEDEVRODEBTAK--ETTEG----- 874
Db 836 AGPGAPSETGETTLEDEEEDEEEDGAEVEGETVDTDEGETVEYQPVKDTRE 895
Qy 875 -----SATTTTSLOVCPYGVKVLTKDNESLDDAGSLKYG--GNNSRL--GWRVTSGE 925
Db 896 GEEBEAKKATDTTTSLOVCDVKNALT--NNDNLTDACKLKGPGKGRFPWKKVVS--SGE 953
Qy 926 -----PTSSDKNGAICVPPRRRLYI---KKIVDATTESPOASGSSEASTGS 972
Db 954 KSVATAGSSAGTGSCKGKALCVPPRRRLYVGLTLKTSAGTSSESPO--CGSSSNASD 1012
Qy 973 -----STPPDSKALLKAFVESALETFFLMHRRKEE---KKAVAQE-----G 1013
Db 1013 VSQNGSGDDITTT-----ESLRKWFIEATAIETFFLMHRRKKEWQAOKKELQNGLLIG 1067
Qy 1014 AGHGLPRVEBSPREYDEBDKL--RKGKIPDGLRQMFYTLGDYRIILSGSNDTISVSKDT 1072
Db 1068 TGAAL--NLGGDSDSNPOTQLOKSGTIPDLFLRLMFYTLGDIYDLIVGVAD----- 1116
```


0Y	1073	PSSNDMLKNTVLLASGTSOEBEKNMKYK--IKNFKRS--TERSPNLVSHPOWME	1128
0Y	1117	---DKNGGNNTLILNAGN--KDEKQKMEKIOEKIEQILPTSGNKETRGONSINROSLMD	1172
0Y	1129	NGGYIHMGWCAITSK--DKIAKGEKKPKIENPENIMDANKPKRPPOYOTPNKLD	1186
Db	1173	RIAEVWHGHVACALITYKDDDNGLKGVKKYFKQIENPEKIMNTTTPKPKDEKTYQOTAKLE	1232
0Y	1187	ENSGTSBPTQTOAASDNTPTTLTJHFKRPPTYFRMEWGESFCERKRLKOITRVDCVY	1246
Db	1233	DESG-----EKRPDSASGCTKLTDKIRBPFRYLEWGEWENFCAKRTMELGKIKEDC-Y	1285
0Y	1247	ENGVDGSCSDGEACDSITHDYSTYPSFNCRCQCKHCSIRKMYLERKKIEPHKQSNAYG	1306
Db	1286	KNG--GRCSDGKLKMEIYIDKEKIRGDLICPTCARHCFYKMKINTKREDFEONSNAYS	1343
0Y	1307	QOK-----TDATRNNGNTFDEKFEKTLTJPMDDAKFLERLKNKPCKTNE--YGGSD	1366
Db	1344	EOKKRYEENDSAOKNNG-----YCGTLK--DDAEFLNRLKKNPCNSESSEENKKADE	1395
0Y	1357	IDFKDSKTFOHTEYCGPCPKFNTCONGNCVSG--LNGNCGDGSIDAKETIARRSSTT	1415
Db	1396	IDFKKPDFTFRADNCKPCSEFEXIKCEHNHNSGGNTQCKQCKTTIATLEBNIKNTYK	1455
0Y	1416	DYWRVSDNDNTNTEBGDLKDACOHANIFKIRKDYWKCGYGVDC--EQTINERT	1472
Db	1456	EVTMLVSDSKSATSEFEDGISECKDKIGIKGIRKDEMECGKGVDCIOMLKKNKNGES	1515
0Y	1473	DGKEYIOIRLFRWVNFLEJDNKINDKISHCIRKKGESKICNGCEKNSCKLEKMEIK	1532
Db	1516	D-KYITIMKELKMWELFYLEDYENKIKHISHCTKNGKSKCIK-----CVRKWQOK	1568
0Y	1533	IAENENIKRRNDQYEKKDDPDY--ANKSILEELPIPIAVNODQVIRKICAFVENSCKTL	1591
Db	1569	KEEKQIKERNEBQYKSKTSDEYFVWKSPLETWIPIPIAVNODQVIRKISFNGSCGSA	1628
0Y	1592	ISNTQNNKENDAIIDCMKLKIGVAKNCKPGKPSGKOSDCK---PPLPDEED--QNP	1645
Db	1629	SAISTGNEEDAIIDCMIKLEKKIDCEKRRKPGNSGOTCNETLTHPLVOEDDEPLEETE	1688
0Y	1646	ENTL--EBPKRCPTTQPREKGEKFCGNKEEKDKKESEEPAKBESSGAABEPATA	1703
Db	1689	ENPVGKQHSPSCPVEDKKKEEGEFC-----TPASPA--	1723
0Y	1704	ESEETENFPEPPCQGAAPSPAPPTPEPLRPQADEPFDSTIQOTTIPFEGVAL	1763
Db	1724	-----PAPAPASPTTAP-----ADEPFDTIQOTTIPPIGIALAL	1759
0Y	1764	GSIAFLFLK-----KTKRASYGNLFQIILPKSDYDIPFLKSSNR	1803
Db	1760	GSIAFLFLKVIYICVVMYIYMCFCIYMKYKTKHPY--DLFSYINIPKSDYDIPFLKSPNR	1818
0Y	1804	YIPIYVSDRIKGYTIYHEGSDDEDKAFMSDITTDY--SESEYEELDINDIYVGSPPYK	1862
Db	1819	YIPIYSGYRKRIYILEGSGTDS--GYTDHYSDIATSSSESEYEEMDINDIYVPSPPYK	1877
0Y	1863	TLIEVYLEPSCNNTAGSKWTSPDSTRNDONCIPESKTIIDEMQMOLKEFISNMLQNP	1922
Db	1878	TLIEVYLEPSCNNTA-----SDTQNDONCIPESKSDNEMKTLKDQFTSNMLQNP	1931
0Y	1923	NDVPNDYSGNSSTNTITTTSHRNVNDNNTTMSRDHNEENMLDPSIHGNCLYSGEYS	1982
Db	1932	KDVPNDYKSGDIPNTQ-----PNTLYDRKREKRPITSIHDRNMLNGEYS	1978
0Y	1993	YVWVM--VNSAMDPIINDNNVYSGIOLINDSLSGKRPDIYDEVYKRRKNEELFTEJENTKR	2041
Db	1979	YVNMKSTKMSMD--PKYVSNVYSGIDLINDSLSGKNHDIYDEVYKRRKNEELFTEJNYKH	2037
0Y	2042	TSTQVAAATTSQDIHNOLEFKMYLDRHDMCEKKKNEDILUNKLEMKMNENNSG	2101
Db	2038	TSIHSYAKNTSPDILQNIOLPFTWIDRHRDMCEKKNENHEKRLAKLEBW--ENETJHSGN	2095
0Y	2102	TYNSDNKRSHNVYLNMTDVSIOIDMDNPKRKNETINNDTQDQSTYDITLIDLEK--YNDPY	2160

ID	AC	Q26031	PRELIMINARY;	PR7;	3078 AA.
Db	2096	THPSOS----	NKTLNFTVDSIQIHMDNPKPIQFN-----	-DITLLEDDLEKFNPEPY	2142
Qy	2161	YDFEFDLIIYDDEVKSSMDIYVDH--NNVTSNNMDVPTKMIENNIANKKEIPE			2218
Db	2143	YDMDKDD--IYDVN-----	-DHDISTVDINADVPKQVEND-VNFK--	-LVKE	2187
Qy	2219	EYPSIDWNI	2228		
Db	2188	KYPIADVMDI	2197		
RESULT	4				
Q26031	Q26031	PRELIMINARY;	PR7;	3078 AA.	
AC	Q26031;				
DT	01-NOV-1996 (TREMBLrel. 01, Created)				
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)				
DT	01-JUN-2000 (TREMBLrel. 14, Last annotation update)				
DE	VARIANT-SPECIFIC SURFACE PROTEIN.				
GN	VAR-1.				
OS	Plasmodium falciparum.				
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.				
NCBI	NCBI_TaxID=56833;				
RX	SEQUENCE FROM N.A.				
RP	STRAIN-DD2;				
RC	MEDLINE=95330813; Pubmed=7606788;				
RA	Su X.Z., Heatwole V.M., Wertheimer S.P., Guinet F., Herfeldt J.A.,				
RA	Peterson D.S., Ravetch J.A., Wellens T.B.;				
RT	"The large diverse gene family var encodes proteins involved in				
RT	cytoadherence and antigenic variation of Plasmodium falciparum-				
RT	infected erythrocytes.";				
RL	Cell 82:89-100(1995).				
DR	EMBL: L40608; AAA75396.1; -				
SQ	SEQUENCE 3078 AA; 349299 MM; C8037C2BC3CCD7C3 CRC14;				
Query Match	40.5%; Score 4901; DB 5; Leng:h 3078;				
Best Local Similarity	36.6%; Pred. No. 5.6e-259;				
Matches 1174; Conservative 277; Mismatches 631; Indels 1128; Gaps	83;				
Qy	4	SGSGGTODEDAKHAVLDEFGQKRVHDEVHGKAKNYVSELKGSLSIASILJETAFTVKSMT	63		
Db	12	AAGGDIDEDSAKHFMDIRIGKDYDKVEAKEKRGKLGGRLS-----EAKFEKNESDP	65		
Qy	64	ES-----KYTELIAENSKRNKC-----KDGKGNVDRRSVKEQAGYDKKKKKC	107		
Db	66	QTPEDPCDLDRHXYHNVTYN-VINPCADRSDFRFSDEYGGQCTHNRIKDSOOGDKG--	121		
Qy	108	SGMTCAPPRRLHLGNKKNPNMNSNDSSAKADLLAETCMAMKYGESIKTHPYKYSKY	167		
Db	122	---ACAPPRRLHVCDDNLEQIEPIKINT--NLLVDVCMAMKFGGOSTIDQPRYQATY	176		
Qy	168	PGSDPMTMLARSPADIGDIIIRGDLVGNKKKQNGKETEREKLQKLEIFKKIDN	227		
Db	177	GPSQICOTMLARSPADIGDIYRGDLVLYGNQELK-----QROOLENKLTIFFGIYEK	231		
Qy	228	LKDKAOKRYNGDEDPNFTYKRLLEDWMTANRETVWGMATCSKLDNSSFPRACNTGQGP	287		
Db	233	INGAEA--RYG--NDPEFFKLREDMWTANRETVWKAITCNAM--GNTYFHAHCN--RG-	281		
Qy	288	SOITHHKCGDRDCKKAGNAGPKRAGDDVITVTPYFYVYQYLWPFEMWEDPCRKKKKLE	347		
Db	282	ETKGYCRNDQ-----VPYFYVYQYLWPFEMWEDPCRKKKKIK	325		
Qy	348	NLEKCRGKDKSDSEYRCGRNGYDEQGISRKGVKVMKGGCTDCEFAQGSYENMTDNQK	407		
Db	326	DVKRNGRGKDKSDXDRYCSRNNGYDEKTRKRAIGKRLRYGKQICISCLYACINPYVDWLNQKE	385		
Qy	408	QFDKO-KYTKTK-----SDGGKKKRAVGC--TTKYEGYKSYELILKNDGYTVAAF	458		
Db	386	QFDKOKKYKDEIKYENGASGSGKQKRDAGGCTTTNTNDGEKKFYDILNNSSEYTFYVF	445		

Qy	4459	IGLLNNERKACADITD--GKINFEKVNSSGGVYGGSGGTSGASGTNDENKGTFRSEYC	516
Db	4446	LEKLSNEITCTKVADEBGGTIDFKNVN-----SDSTSGASGTNVNDSGTFRSKYC	496
Qy	5117	QPCRCQOQ--HKGG--NOMERTKYKKNMKSLYPR---TGNKVLLSLKYVKDMIT	569
Db	4497	QPCRCQKKNVNNSSGSSNENMEKKN-GCKSCGLEYPRKDKGTYITLLKSGKHDD--	552
Qy	5710	LKKNNKEFCLTQNSSDSVSVSYVTGTGASGNSKELTYDEMCKYKHNEQKVNVOGEVE	629
Db	553	IEEKLNKFCDEKNDDTINSGSGSGTGGSGGSGNSROELYEMCKCYGEDVYKXGHDEDE	612
Qy	6310	DDDELKAGGGLCIIIPNPKKKEVSEAKSONNHADIKTEHDFEYVVAHMLKDSITHWTK	689
Db	613	DYENKKNAGGGLCIILNOKKKKEEGSGTSEKEPEIDQITNPFYFVVAHMLKDSITHMK-K	671
Qy	6910	RLKSGISGKTMKC--RNGCKKGDCCPEKWKQKQETEMKIKXHFQTOBEPGRYFTLE	748
Db	672	KLQRLQNGNRKIKGCKNNKCNNDCECFKRWITQKQKDMGKITVHFPTQNTKRGSGDNTAE	731
Qy	749	LI-----LKLOFLKED-----TEENTENSLDAEAEELKHLOILKLENNENLAV	793
Db	732	LIPFDHVLQYNLQDEEFLKGDSDASEKSEKSLDAEAEELKHILREISEDNQNAS	791
Qy	794	VNAG--TEQKTLMDKLLHLELNATKCKOCPLEPDEK-----	828
Db	792	VGGGVTEQKNTMDKLLNVEKDEADLCLEIHEDEEKEKGDGNECIEEGENRNPSCGE	851
Qy	829	-----	828
Db	852	SGNKRYPLAKKVAQYOHNNHAKTQOLASRAGSALRBDISLQAPKGRNGSITLKQOICKIN	911
Qy	829	-----SKRSADPS-----PDIFLP-----	843
Db	912	ENYSNDSRGSGGPGCTKQDGHGVMRICTEWSNIEGKQTSYKNVFLPRPREHMTSN	971
Qy	844	-----	843
Db	972	LENLDVGSVTKNDKASHSLGVDYQLAAKTDAEITIKRYKDNQNIQLTDPQKQDEAMCR	1031
Qy	844	-----	843
Db	1032	AVRYSFADLGIIINGRQMBEDKSSYTDMEIRLITVCKNIKKEHNDGIXDNPKYTGDESKP	1091
Qy	844	-----	843
Db	1092	AYKRLRADWMEANRHOVMRAMKCATKGIIICRPMVPDYIIPRLRMWTEMAEWYCKAQSQE	1151
Qy	844	-----	843
Db	1152	YDKLLKICADCMKSGDKCTQGDVDCGCKCAACDKYKEIEIKEMQWRRKISDKYNLYLQ	1211
Qy	844	-----	843
Db	1212	AKTSTNGPRTVLDDDDPDYQOWVDELTPRIHKASIAARVLVYKRAAGSPETIAAARITPY	1271
Qy	844	-----	843
Db	1272	STAAGYIHOEIGYGCQEOQTQCEKKHGAITSTTEKNEYTFKOPPEYATACDINRS	1331
Qy	844	-----PREEKED-----	850
Db	1332	QTEBPKKKEEVNESAKEYIKIIEGKNKRTTYGECNPKSESYDMDCKNNIDISHGACMP	1391
Qy	851	-----DEN-----BDDDEDEYRD-----	863
Db	1392	PRQKLCUYVIAHESQENIKTIDNLKDAFIKTAAAEFFLSQYKKSNNDBSAKILDRGL	1451
Qy	864	-----DEBTAK-----TTEGS-----	875
Db	1452	IPSOFLRMMVTFEDYRDICLNTDISKKONDVAKADKIGKFFKSGDKSGKSPGLSQROEMW	1511

QY	876	----	-ADT-----	111	-----	879
Db	1512	KTNGEIKMGICALTRYVDTDONKRIKNDYSYDRVNOSONGNSPLEEFAKPOFLRM				1571
QY	880	-----	-----	-----	-----	879
Db	1572	IEWGEFCAEOKKENIICACNEINSTOOCNADKIRHCQACRAVOEYENKKKEFSOT				1631
QY	880	-----	-----	-----	-----	879
Db	1632	NNFVLKANVQPODEYKGYEKDGVQIQGNEYLLQKDONNKCSCMDGNVLSVPEKPF				1691
QY	880	-----	-----	-----	-----	903
Db	1692	GKYAHKYPEKCDYOGKHVPSIPRPPPVYOPOEAPVYVDCSTV-KTLFDMNNFSDA				1750
QY	904	CSLAKYGVNNSRLGRCY-----TPSGEPT-----SSDKGALCVPPRRRLYIKKIYMAWTK				957
Db	1751	CGLKY-GKTAASSMKCLIPSDTKSAGATTCKSSDSGSLICIPPRRRRLYVGLQEMATL				1809
QY	958	ESPOASGSEASSTGSGTTPPDSKEALKAFAVESAAIEFFLHRRYKEEKKAVAGAGHG				1017
Db	1810	--POGEAPSHSRA-----DOLRNATIQSAALIEFFLMDRYKEEKKPPG-DGSOQA				1858
QY	1018	LPRVEE-GSPYEPEK-LKEGIPDGFRLQMFYTLGDRILHSGSDTTSYSKOPS				1074
Db	1859	LSQLTSTYSDDEEDPPPKLLQNGRIIPDFRLMFYTLGDRILHVGGN--TSDSGNTNG				1916
QY	1075	SSNOLKIVYLLAGSTGEOREKMNKKKEI--KNFKCSGSEASAPLVSHPOTWMENNGK				1132
Db	1917	SNN--NIVLEASGNKEDMOKIOEKIEQLPLKNGSTPLVPKSS--AOTPKMNEHAE				1970
QY	1133	YIWHGMVCAIT---SKOKIAKVEKKPOKIEPNELAMDE---ANKK-----BKP				1179
Db	1971	SIMGMICALITYTEKENPDTSGARDE--NKIEKDEVEYKEPFGSTADKKHASTPTGYTK				2027
QY	1176	POYOTYVNLKENDSGTBPRTTOTOASDNTPTTLTHFYKRPITYFRNFEKSGSFCREKK				1235
Db	2028	TOYVEYKEKLEDTSG---AKTPASSDTP-LISDFVLRPFYRLYELMGONFCFKRKH				2081
QY	1236	RLKIOIKVDCKEVENDVG-----RCSGGEACDSISHTDYSIVPSFNCPCGCKHGSSY				1287
Db	2082	KLAQIKHECAVEENGSGSRGGITRQISGDGEACNEMLRKNDGTVPDLKPBSCAKPCSSY				2144
QY	1288	RKWIERKKIEFHKOSSNAYGOOKTDATRNNGNPTFDEKFCCTLETPWDAFLERLANGCK				1347
Db	2142	RKWIESKGKEPEKEKAYEDQK-DKCVNGSNKHDNGFCETLITSSAKOFLKTL--GPKC				2198
QY	1348	TNKEYGDDIDFEKDSFTPOHTEYCGPCKPFTKNGCONGCYSGLNGCNDGKSIDAKEI				1407
Db	2199	PNNVEGKTIED--DDKTFHNTKDCBCLKFSYVCKKDECD--NSKOTDCRNKNSIDADI				2254
QY	1408	AKMSSITTDVMRVSDDTTFEGDDOLKDAJCOANIFKGIKRDVMYMGCVGVCVDICEQTN				1467
Db	2255	ENGVDSTV-LEMRASASKSGFNGDGENACRAGIFBGRIDEMKCRNVCYIVCAKEN				2313
QY	1468	INERTDKEYIOLALEKRMVNFLEJDNINDKISHCJIKGEGSKCJINGCEKNSKJLEK				1527
Db	2314	VNGEAKGHIIQIRALVKRVNVEYFFEDYNNKIKKISHIRKINEISPCI---KN-CVEK				2367
QY	1528	WIEKIAEWENIKKRENDYQENKQDPDYNVKSLIEBLDKIAVVNDODNYIKLCAVENS				1587
Db	2428	GCSASANQKNGEYKQADICMLKRLKIGCEKKNHOTSTESDTPROPTLEDFLD				2487
QY	1640	--BDONPEENTLEPPFCPT--TORPEKGETGSKNEEKKDEKSESEPAKKEESGP				1694
Db	2488	DDITEBAKKMM--FKICENYKLTQAOEDBG--C-----VPAENSEPAATDSGK				2535
QY	1695	AAEPAPTAASEETENTFPPEPCTGPAAPSTPA--PTDTPPLRLPADDEPDSITL				1751

Db 2536 EPEGTPLVKPEEAVPEPP-----PROKAPARIPQPPRPPOLDNPHVLT 2589
Qy 1732 QT-TTPEGVALLGSIATLEFLKKTAKSYGNLPOLLQIPKSYDIPTLKSSNRYIPYSD 1810
Db 2590 VSTLAWSGVIGFAFTFYFLKTKTSSYGNLFQIOLIPKSDYDIPTKSPNRYIPYSG 2649
Qy 1811 RKKGTYIYMEGSDDEDYAFMSDTDTSSSEYEELINDIYVPGSKYTLLEVYLE 1870
Db 2650 KRGKRYIYLBGSDGTDG- GTYDHYSDITSSSEYEEMDINDIYVPGSKYTLLEVYLE 2708
Qy 1871 P-----SGNNTTASGKNTPSDTRNDIQNDGIPSKITDNEMLKKEFISN 1916
Db 2709 PGNNTTASGNNTTASGNNTTASGNNTPSDITQNDIQNDGIPSKITDNEMLKKEFISQ 2768
Qy 1917 MIONOPNDVPNDYSGNSSTNTNTTTSRHVNDNTTMSRDNMEENILPSIHGONLY 1976
Db 2769 YLQSEPTNEPN-----MLGYNDNNTHTPTSHHNEEKPFIMSIHRLNF 2813
Qy 1977 SGEESYNY-----MMVN-----SMNDIPINDNNVYSGIDLINSLSGKPI 2019
Db 2814 SGEETNYMFPNSGNPINSSTNSMDSLTSSNHSPLYNDKNDLYSGIDLINDALSGNH-I 2872
Qy 2020 DIYDEVLRKKELELGTGTE-NTKRTSTQNYAKTNSDPINHOLELPHKMLDRHRDCKEKK 2078
Db 2873 DIYDEMLRKKELELGTGTHHTKHTNTYNAKAPARDDPITNOINTLHKMLDRHRDCKEKK 2932
Qy 2079 NKEDILNKLKEEMKNENINSKTYNSDNKPSHNHVLNTDVSIOIDMDNPKTKNETND 2138
Db 2933 NNHERLPLKLEIM--ENETHSGDI--NSGIPSGNHVLNTDVSIOIDMDNPKTKNETND 2988
Qy 2139 TNOGSTMDTIDDLLEKNDPYYPYEDDITTHDNYDEKSSMDIYDHNVTSSNMDY 2198
Db 2989 TNPDSSTMDTIDDLLEKNEPYYPYEDDITTHDNYDEKSSMDIYDHNVTSSNMDY 3048
Qy 2199 PTKMHIEMNIVNKKKEIFESEYPISDIWN 2228
Db 3049 PTKMHIEMNIVNKKKEIFESEYPISDIWN 3078
RESULT 5
Q26034 026034 PRELIMINARY: PRT: 2182 AA.
AC 026034:
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)
DE VARIANT-SPECIFIC SURFACE PROTEIN.
GN VAR-7.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DD2;
RX MEDLINE=95330813; PubMed=760678;
RA Su X.Z., Heatwole V.M., Wertheimer S.P., Guinet F., Herrfeldt J.A.,
RA Peterson D.S., Raveloch J.A., Wellem T.E.;
RA "The large diverse gene family var encodes proteins involved in
RT cytoadherence and antigenic variation of Plasmodium falciparum-
RT infected erythrocytes."
RL Cell 82:89-100(1995).
DR EMBL; L42636; AAA7539.1; -;
SO SEQUENCE 2182 AA; 248722 MW; 813969A01460B4F8 CRC64;

Query Match 39.8%; Score 4817; DB 5; Length 2182;
Best Local Similarity 45.9%; Freq. No. 1.4e-234;
Matches 1096; Conservative 298; Mismatches 607; Indels 388; Gaps 93;

Qy 4 SGGSG---GTDP-----EDAKHYLDEFGQVHDE--VHGAKNYVSELKGSLSLAS-I 50
Db 11 SGGSSSGGKKKDTSEIYVSDAKDLDRGVEKVEYBEKYNKGAKKIIEALKNLNTANGR 70

Qy 51 LGETAFTVSKMOTBESKYTELLEIANSKRNPCRKDKGNDVRS-----VKEQAG 99
Db 71 SSETASSTETCTLVKEYERYVGDGRHPCRDANEDVNRPSDTLGO:TYNRIKIDSOO 130
Qy 100 YNNKMKKCSNGMTCAPFRLHLICNNKFPNMSNDSKAKHDILAECYMAIKYEGESIKRH 159
Db 131 GNNKY-----GACAPYRLHLCDY--NLBSIDTSTHKLILLEYCAAKDEGNSINRH 181
Qy 160 YPKYDSKYPGSDPFMYCTMLARSPADIGDIIKRDLYLG--NKKKQNK ETEREKLQKL 217
Db 182 YHOHRTNEDSASQCTYLARSFADIGDIVRGKDLYLGYDNKEKQ-----RKLEQL 235
Qy 218 KLEIFKIH-DNLKDEAKQRYNGD-EDPNEYLRDPMANETWGMATCSKEIDNSY 275
Db 236 KDIFFKIHDKVKTGNAQRYIDDAKGGFFQLREDMWTSNREYKALICHAPREANYE 295
Qy 276 PRATGNDTGQSPQSTHNKRCDCDKGANAGKPKAGGDVTVIPYEDVYPOYLRFEEWA 335
Db 296 IKTACH-VKRG---INGQCHC-----IGGD-----VPTFYDYPOYLRFEEWA 335
Qy 336 EDFCRKKKKLENLKQCKDKSDERYRCSBNYDCBOTISRKGKVMGKCTDQFPAC 395
Db 336 EDFCRKKKKLENLKQCKRDYEQN---LYCSGNGYDCOTIYKKGKLYIGCHCTNSYWC 392
Qy 396 GSEYMWIDNQRKQFQKQ-KYTKETISDGG-----RKKRAVGGITRYE--GYEKSFEKL 447
Db 393 RMYETWIDNQRKEFLKQKRYETETISGGSGSPKRTKTAARSSSDI NGYESFEYKL 452
Qy 448 KNDGYTVDAFGLLNNEKACADITDGGKINFEVNSGGVYGGSGGI SGASGTNDENK 507
Db 453 KEVGYDQVDFLKLKKEICQKQPOVG--NEKADN-----VDFTNEKIV 495
Qy 508 GTFYSEYQCPQDQVQHKGNGQWERKTKYKMRKSKLYKINGKMYL-LKSLKVVXKD 567
Db 496 KFEFSREICEPCPWCGLG-KGGPPW--KVGDKTCGSAATKYTDPRNIDIVLPDKSQ 552
Qy 568 MILKKNMKEFCLTQNSSDGVSQVYTTGASGNSSEKELDYEMKCYKHIEVOYKVNQGEV 627
Db 553 ONILKRYKNFC-----EKGAPGGGQIKR-----WQCY----- 580
Qy 628 EDDDELKAGAGLCTLPNPKKKEVSEAK-----SONNHADIQKTFHDFIYYVVAHLKDS 683
Db 581 ----DEHR-----PSSKNNNVCYEGTWKFTQCKOT--VKSIVFVWIDVHDLHDS 626
Qy 684 IHMPTKRLKSCI---SDGKTMKCRNGCNKKCDCEFKVYKQKETEMKPIIDHFKTOGI- 738
Db 627 VEMKTE-LSKCLNNNTNGMTGNKNNCKCTDGCQFOKWEKCKQOEWMAI:IDHFGKQTDIVQ 685
Qy 739 -----PEGYFTTLELILK----LQFLKEDTEBENTENSIDAEBAE LKHLQIKLLE 786
Db 686 QKGLIVFSPYG---VLDDVLKGNLQNIK-DVHGDV-----DJIKHIKKL--LD 729
Qy 787 NENNLAVVNAAGTEOKTMDKTLNHELNDATKCK---DCPLRDEKSR RSDSPSDIFI 842
Db 730 EEDAAVAVVLGGKDNFTI-DKLLQHEKEQAEQKQOEBCERKAQDSR RSDAETREDEBT 788
Qy 843 PPP-----EEKEDENEDDEVDDEETAKETTESGATPTTSLDV----- 885
Db 789 QGPASAGVEBEDDDDYDDEDDDDVQOEBEKE--EGVITEVTEVTEVVEEIVTQ 846
Qy 886 ----CPVIGKVLTKDNESLDQACSLKY--GGNNSRLGNRCVTPSGEPTTSSDKNGAICV 938
Db 847 EGVKPCDVIYK-LFEDDKSLKACGLKYPGKKEKFPNPKCVTPSGVSPATSGKGAICV 905
Qy 939 PPRRRRLYIKKIVDMATKTESPOASGSEASTSGSTTPPDSK---EALIKAVESAITP 995
Db 906 PPRRRRLYVGGISQWASR-----GGDETTEVSSSEATSPASQSESEKLRTAIESALET 959
Qy 996 FFLMHRKYEEKKAVNA-OEGGAGHLPRVEGSPEDYEDPELKL-GKPLGFLQMQEYTLGD 1053
Db 960 FFLMHRKYEEKKPRATQDQAGLGSVLPSPSPGDEPQDQLQGTGVIPFDFLQMQEYTLGD 1019

Qy	1054	YROLLESGSNDTT--SVSKDPPSSNDNKLKIVLVLASGSTEOREKEMN-----KYKEIKNFR	1108
Dd	1020	YKDLVSGSSNDTSPOTSTOKTPSPSSNDNKLKIVLVLASGSTEOREKEMNIOQAIRKILMGA	1079
Qy	1109	KCSTERSAPNLVSHPR--QTMWENNGKXYIMHOMCALMSKDKIAKGYEKK--POKIENTENIM	1166
Dd	1080	TSQVPRPTKNSVKKPPOOTWENENIANKLOWMANMAYCALLYKRENDARKGSAKIEONKDLKRLYM	1139
Qy	1167	DEANKKRPPOYQYTNVKKLDENSGTSPPRTQTOASSDNT--PRTLHPRKPYTFYRFE	1224
Dd	1140	DEANKNPPIEYQYTNVKKLDESG-----AKSNDTIOPPTLKNEVEIPTFRWLMH	1190
Qy	1225	MGESCRERKRLKQIYNODCKVENGVDGRCSGDGEACDSISNDHSTYSPSFCNPGCCGHC	1284
Dd	1191	WGNSTCERARRLQIHLHECMBEDGE--KQYSGGCEEFISFOQYNVLQDLS--SSCAPCP	1248
Qy	1285	SSYRWKIERKKIEFHKOSMAYGOOKTD-----ATERNONTPEFKDEFCKTLFTRPDA	1335
Dd	1249	RLYTWIEKKTEYKQKQKAYBQCKSNYEKQDKQCTOSNMNA--NEFSRLGASPTPA	1306
Qy	1336	KFLERLKNGPCITKNEX--GGDD--IDFEKDSKTFQTEYGCSPKFKTNQCNCGVSGL	1392
Dd	1307	EFLQKRL--GSCCKNDNGENEDENKIDFKNPDKTFKAHSCDCPITGVKQCNQHC--VESA	1363
Qy	1393	NG--NCDDGKSTDAEIKAMMSSTTDVVMYNSDNTFRF--GDDLKDACQAHNFKGRK	1449
Dd	1364	NGKECKKNK--TTAEDIKNKTDPNGNIEMVVSDSTTFEHLDD-----CKSGSIEGIRK	1417
Qy	1450	DVMKGYVCYGVYDLC--EQTMINERTGKEXEYIOALFKFWEENFLFEDYKINDKRISHIK	1507
Dd	1418	DEMGCANVCGYDICTLEKKIKNOBQDCKITYKELKKLWMEYFLFEDYINRIKRIKILCTK	1477
Qy	1508	KGBSCKINCCEKNSKCLEKWIEMIKIAEWENIKKRFENDQYENKDDOPYNVASILEELPK	1567
Dd	1478	KEDCKCKIKG-----CIEKMWQEKTEKMOKINDTYLEQYKND-----GNLTNLEQ	1525
Qy	1568	IAYVNDODNYKLC-----VEENSGCTLSNPNKNENADIDAMOKLKLGVAKNCPGRS	1623
Dd	1536	FQYTEREKNNAIKPCDGLDOKFTKSCGLNSTDNGSNGNNDLVLCJLKLKIKISECKEHS	1585
Qy	1624	GEKOSDC-----KEP-----PPLPDEEDONPEENTLEPFCPPPTPOPEEKGCE--TCG	1671
Dd	1586	GQOTPCDNSLSKKESTLYVEDDYDEBQNP--ENKYEQKFCFPMKKEPKKENDDEVGCG	1644
Qy	1672	NKEEKDKEESEEPAKEESGPAABEPAP--TAESSETETN--FEPBPGTGAAPSTPA	1728
Dd	1645	GDEKK--KVEDSYIEQKEEELASABEESPPLTPEAPKKEENVPKP-----	1690
Qy	1729	PPPTDPTP-----LAPQADEP--FPOSTIQTITPFGVALALGSAFLFKKTKRASGN	1781
Dd	1691	-----PPPKRKRIKTYNNVLNDHVAJVALMSSTIMWISIGFAAFTYFYLKAKKSSYGN	1744
Qy	1782	LFOILQIPKSPDYDIPTLKSSNRYIPIVSDRYGKQTYIYEGSD--EDUYAFESPTDVT	1840
Dd	1745	LFOILQIPKSPDYDIPTLKSSNRYIPIASDRHKQTYIYEGSSDDEKXAFNSDTTDT	1804
Qy	1841	SESEYELDINDIYVPGSPKYKTLLEVLYLEPSGNNTTASGKNTPSDTRNDIQND--GIP	1897
Dd	1805	SESEYELDINDIYVPGSPKYKTLLEVLYLEPS-----KRDQNDINDIPSDIP	1853
Qy	1898	SS-----KITDENNOUKKETISMALONQRPNDYDTSGSNSSTNTNTTSSHNVNDNTN	1933
Dd	1854	NSDPRPPTIDENNOUKKIDISMALONTONTPEP-----ILHDVNDNTH	1898
Qy	1954	TTMSRDNWNEENLLPSTHODNLYSGEYSXV-----NNVN-----SMNDIP	1996
Dd	1899	PTMSRHNMDOKPFLMSIHDNLLSGEELNTDMENSGNPNINISDSTNSMDSTJNSNHSY	1958
Qy	1997	NRDNNVYSIDLLINDLSGKRPDIYDEVALKRENEFTEWTKRTSTQN--VAKTINS	2054
Dd	1959	NKDKNDLYSGIDLLINDALSNGH--IDYDEMLKRENDLROTCQHNPKNITSNKNVYQJSSD	2017
Qy	2055	PIHQBLEFHKWLDRHDMCEKKMKNEIDLNLKLEBWKENINNSGKTYNSDNKPSHNH	2114

ID	Accession	Score	Length	DB	Score	Length	DB
Db	2018 PTN001NLFHKKWLDHRDCEKAKNNHEPLKXELM--ENETHSDI--NSGIDSGNHV	39.7%	4809.5	DB 5	38.7%	4762.5	DB 5
Qy	2115 LNTDVSIOQIDMNDPKTKNEITNMMDTNDQSTYMDTITLDDLEKYNPDYFDEDDIYHDV	38.7%	4762.5	DB 5	38.7%	4762.5	DB 5
Db	2074 LNTDVSIOQIDMNDPKTKNEITNMMDTNDQSTYMDTITLDDLEKYNPDYFDEDDIYHDV	38.7%	4762.5	DB 5	38.7%	4762.5	DB 5
Qy	2175 DVEKSSMDIYVDHNNVTSSNNQDPTFKHIEKNYNNKKEIFEDEYPIIS 2223	38.7%	4762.5	DB 5	38.7%	4762.5	DB 5
Db	2133 NDDKASEDHIMDKMNNNSDVPETNVOIEKNNVNN--OELLQNEYPIS 2180	38.7%	4762.5	DB 5	38.7%	4762.5	DB 5
RESULT	6	PRELIMINARY	2710 AA				
Q9XZB8	Q9XZB8	PRELIMINARY	2710 AA				
AC	Q9XZB8	PRELIMINARY	2710 AA				
DT	01-NOV-1999 (TrEMBLrel. 12, Created)						
DT	01-NOV-1999 (TrEMBLrel. 12, Last sequence update)						
DE	01-MAY-2000 (TrEMBLrel. 13, Last annotation update)						
DR	VARIANT-SPECIFIC SURFACE PROTEIN.						
GN	Plasmodium falciparum.						
OS	Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.						
OC	NCBI_TaxID=5833;						
RN	SEQUENCE FROM N.A.						
RP	STRAIN=CS2;						
RA	MEDLINE=99236507; PubMed=10220443;						
RA	Reeder J.C., Cowman A.F., Davern K.M., Beeson J.G., Thompson J.K.,						
RA	Rogerson S.J., Brown G.V.;						
RT	"The adhesion of Plasmodium falciparum-infected erythrocytes to						
RT	chondroitin sulfate A is mediated by P. falciparum erythrocyte						
RT	membrane protein 1."						
RL	Proc. Natl. Acad. Sci. U.S.A. 96:5198-5202(1999).						
DR	EMBL: AF134154; AAD29126.1; -						
SO	SEQUENCE 2710 AA; 309249 MW; 0AC64FOA3D5BF512 CRC64;						
Query Match	39.7%; Score 4809.5; DB 5; Length 2710;						
Best Local Similarity	38.7%; Pred. No. 4.7e-254;						
Matches 1109;	Conservative 308; Mismatches 646; Indels 801; Gaps 80;						
Qy	5 GSGSGTQEDDAKHVLDDEFGQKHYHDEHGCAKIVYSELKGSLSLAILGETATVVSMTQTE 64						
Db	8 GGSSEDDDKAKHVLDSIGEKVYKVKVAAYSSQLKGLTSLNAFENEDP---KGQOYA 63						
Qy	65 SK-----YTELLEANSKRNPCKKQKGDVDFRSYK 95						
Db	64 NDPCLLLEYHFNNTYKGGHREHPCKLLLEYHTNTYTKGGHREHPCK---KGLE-KRFSYV 119						
Qy	96 EDQAYDNKKMKCS--NGMTCAFRRLHLCKNFPNMNSNDSSKAHND--LAEVCAAKYE 152						
Db	120 GGGCEDDNKRKIKOSKNNNGACAPYRRLHLCVRLLENISALD--KINNDTLADVCLALHE 177						
Qy	153 GESITTHPKYDSKIPGSDFP--MCTMLARSPADIGDIRGDLVLGNKKKKQNGKEIER 210						
Db	178 GGSITQDPKPYQAOYASSFPSQITMLARSPADIGDIRGDLVLGNKKKKQNGKEIER 231						
Qy	211 EKLEOKLEIFPKIHNDLNDKDEAKORRYNGDEDPNPFYKLLREDMTNRRFVGMAMCSKEL 270						
Db	232 --LENNLNKKIKGKIEKTLTDPRAKHQHYKDEDPNNFFQLREDMANNRRQCVMAITICHCAG- 288						
Qy	271 DNSSYFRATCNDTGGGSPQSTHNKRCQDCKDKGANKAPKAGDGVITVPTPYDVPQYLHW 330						
Db	289 ESDKFRKRTACGTG---TQGRGCRNDK-----KPSNTDPPYFVYVQYLHW 336						
Qy	331 FEEMAEDEFCRRKKKKLEMLEKOCRGKDKSDERYCSRNGYDCEQITISRKGVYRMKGCTD 390						
Db	337 FEEMAEDEFCRRKKKKIKDYKRCRG--ERDGEKYSRNGYDCEKTKRAIGKRYMNGCTK 394						
Qy	391 CFACGSGSTENMTDNRKQPDQ--KYTTEISIDSGGRRKRAVAGTGTTKYGYEKSFEKELKN 449						
Db	395 CLFACNPPVEMTDNRKQPDQVYKVEISIDSGGRRKGA--ATIKRYGSEKFEKELKS 452						

[illegible]

Qy	1046	QMF7TLGDYDILFSGSND--TTYSVKDTPSSNDMLKATVILLASGSTEJ2REKANKYKE	1103
Db	1490	LMFVTLGDRIDICGKTDPDGDIDIVSQRDMENIKRAIEVSFPSSGTSPEVETKTR--	1546
Qy	1104	IKNFRKSTERSANVLVSHQOTWEMKNNGYVIMHGWCALTSKDIJAKVEKPKPIEMPE	1163
Db	1547	-----ESMMEKNGEHIMGMICALITYKENGADSTVKIEKDSDYK	1586
Qy	1164	NLMDEANKPKP-----PPQYTNVVKLDENSGTSPPRTJTJASSDNTP--	1207
Db	1587	KIFGNNTDPPDKLITVTYDVKGTFFSKRYDEVELKEDENEGASAS3JLSGD-TPTLN	1645
Qy	1208	--TLTHEVRYKPTFRWFEWGESFCEBKRLKQIKVDCKYENGVGRK3GDGACOSIS	1265
Db	1646	NPKLDEFPKLPPEFRMYLHEGSDFCVYKQAMOLEVADNCLKRDGRTKK3GYGENDCIL	1705
Qy	1266	THDVSVPSEFPCGCGHSGSYBKWLEARKIEPHKOSMAYGOQKTDARNGNENFDEFC	1325
Db	1706	SOKNIEFDEPCPSKRCGLYKRWLSRKRAEDKOSNAYGOQKCY-NNNNHDEFC	1764
Qy	1326	KTLFEPDPAAEFTERLNGECKT-NKEYG---GDIDFESKDSPTQHYEYCGCPKPKTN	1381
Db	1765	RTLK--DDAAQFLENT--GSKKNDNNEBGRKRDILDFKPRQTHQHYDPCDSQFKIN	1820
Qy	1382	CONGCGVSLGNGC-DGD-KSIDAKEIAKMSSTDVVMYVSDNDYTFBEGDLKDAQ	1439
Db	1821	CRNGCKSDDTNSCKDKDKDITAYDKDTPNGENIMLVSDSTGVG3-DLND-CI	1878
Qy	1440	HANIEFKGRMYMKCGVYGVCDICEQTNINERDCKEYIOIALKRHWENFLDYKNIN	1499
Db	1879	KAGIFKGRKAMKCDNVGCGVYVCKPBGKGNOKIITIALLRHWEFTLEDYNNIR	1938
Qy	1500	DKISHCIKGEBSKINGCCKNSKCLEWETKIAEMENIKRFNDQYENKODQPYNKS	1559
Db	1939	KKLNCAIKKONGSTCINGCEKCKOCYQWYTKREEMGKIKDRFNDOYKKNDSGYPYKT	1998
Qy	1560	ILEELIPKIAVVDQDNVYKLCYFENSGCTLSNTQNNKENDA----IDCMKLKLGVK	1614
Db	1999	ILEELIPQIAAATQKGHHKRLQKLVKSLKCNCTDNSEKGEJIDANKDKIDILLDKLKK	2058
Qy	1615	AKNCGKRSKGKODCKEPPRLDEED-QMPEENTILEPPKPCPTTOPPEEK--GGEIC-	1670
Db	2059	TESCPTPTSDHTQACVDSHLEDEEPLLEEENPATOYNICP--QOKFETKVDEDCI	2116
Qy	1671	-----GNKEKDEKKEESEEPKESGPAEEPPAFASE-----EPETN	1711
Db	2117	TDAPDPVKEEBEKEEBKDEDEEEDEEDDEDEEBEESVDTYDUSDETFED	2176
Qy	1712	FPEPPTGPAA-PESTPAPPTPDTPEPLRQADPEPSTLIQTTIPFGVALAGSIAELF	1770
Db	2177	DEDEYVLTSSHSSSQPKRLPRDLSPELK-----KAMELSTILMMVYGJGFAICYFL	2228
Qy	1771	LKKKTKASVGMFOILOTIPKSDVIDPLKSSNRYIPYVSDRKKTIIYIMEDSDSEKYA	1830
Db	2229	LKKKPKSPV-DLLFVLLVHKGDYGTPTPKSSNRYIPYRSGPYKGTIYIMEDSDSGHY-	2286
Qy	1831	FMSDTPVTSSESEYEELINDIYVPSPPYKTLIEVLEPESGNMT-----ASGNT	1883
Db	2287	-YEDTDTITSESEYEELINDIYVPSPPYKTLIEVLEPESGNMTYVSGNTRASGNT	2345
Qy	1884	PSDTPNDNDQNDGIPSSKITDNEMNOLKEPISNMLQNDQNDVPNDYSGNSSTNTNTTT	1943
Db	2346	PSDTPNDNDQNDGIPSSKITDNEMNLTKHDPISNMLQNDQNDVPHDYISGNSSTNTNTTT	2405
Qy	1944	SGHANVDNNTTMSRDNKEERULLPSIHGONLXSGEYSYNNMYNMSKMDIPINRDNV-	2002
Db	2406	SRHNV-----EKPFLMISHDRILYTGEEYVDMSTSGNNDL-YNKKNNLY	2451
Qy	2003	-----YSGIDILNLSGSGFKIDLYDEVLKRR	2029
Db	2452	SGQNNVYSGIDPDSNNGPYSOKNDRISDNHHYTSGLDILNTLSGKNKIDYDEBLKRR	2511
Qy	2030	ENELGTE-NTRKRTSTONVAKTNSDPIHNOLELFKHWLDRRHRDCKEKKNKEDILNKK	2088

```
||||| : ||| : | | ||||| ||||| ||||| : | |||
Db 2512 ENELEGTGHHPRKRTTNHFAPTRDPLHNLNLFHKWLDHRHRCCEKWEHHERLAKLK 2571
QY 2089 EEMKENINNSGKTYNSDNK-----PSHN-----HYLNTDVSIOIDMDN 2127
Db 2572 EEMENDT-STGSGNTHPSGNTPTSDIPSGKOSDIPSDNNHSHDIPYVLTNTDVSIOIHND 2630
QY 2128 PRTKNEITNMDTNODKSTMDTILDLLEK-YNDPYRYDEVEDDIYHDYDVEKSSMDIYV 2186
Db 2631 PKPIQFTN-----MDTILDLDLDPKPFNEPPRYDYDD-YYDVH----- 2669
QY 2187 DHN--NVTSNMNDVPTKHHIEMNIYNNKKEFEEEPYSDIWT 2228
Db 2670 DHDPTVDTNAVDPSKVQIEMD-VNTR--LVKEKYPIADVWDI 2710

RESULT 7
Q26033 PRELIMINARY; PRT; 2664 AA.
AC Q26033;
ID Q26033;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE VARIANT-SPECIFIC SURFACE PROTEIN.
GN VAR-2.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FCR3;
RX MEDLINE=95330813; PubMed=7606788;
RA Su X.Z., Heatwole V.M., Wertheimer S.P., Guinet F., Herrfeldt J.A.,
  Peterson D.S., Ravetch J.A., Wellem T.E.;
  "The large diverse gene family var encodes proteins involved in
  cytoadherence and antigenic variation of Plasmodium falciparum-
  infected erythrocytes."
RT EMBL; L40609; AAA75398.1;
DR CELL 82:89-100(1995).
SQ SEQUENCE 2664 AA; 302412 MW; 6EA2468511703091 CRC64;

Query Match 38.98; Score 4711; DB 5; Length 2664;
Best Local Similarity 39.38; Pred. No. 1.le-248;
Matches 1116; Conservative 290; Mismatches 639; Indels 796; Gaps 88;

QY 2 ATSGSGGTQDEDAKHYLDERGKVDHVEGEA-KNYSELSGLSLASIGETAFTYKS 60
Db 6 AAGGGGSDIDRSKAHLDISTGKRYDVHGDALQPSNGKLGTSLA--TFEKAPEGKQ 63
QY 61 MOTE-----SKYTELEANSKRNPCKKGDNDVDRFSVEQAGYDNKKKCSNGM--- 111
Db 64 TSEDPCLDNHHTYTVJSGYKENPC---KDRPEVRSTYEGAECDKSKIRGNSNKDG 119
QY 112 TCAPFRRLHLCNKFPMNNSNDSSAKHDLAEVCAAKYGESIKTHYPKYDSKYPGSD 171
Db 120 ACAFFRRLHLCDOHLEIKHKDKIT--RHNLADVCEAAKFEASELEKRGQYLNSDYN 177
QY 172 FPMCTMLARSPADIGDIIRGDLTGNKKKKKQNGKTEPRELEOKLKEIFKIDNLDK 231
Db 178 INITELARSPADIGDIYRGDLTGNDK-----EKDLEENLRIFKFIKIDNLDND 229
QY 232 EAQKRYNGDE--DNPFKYLRDMWTANRETVWGMATCSKELDNSSYFRATGNDTGOGSPQ 289
Db 230 HVQHHYHDDDGKTKNYKRLRANWEANRQTYWKAITGCA--GGTYRQTC---GTG-TW 283
QY 290 THNKCRCDDKDGANAGPKRAGDGVTVPTVFDYVPOYLRFWEEMADFCRKKKKLENT 349
Db 284 TNEKRCRC-----PINDVPTVFYDVPOYLRFWEEMADFCRKKKKKYDIY 327
QY 350 EKQGRGDKSDSEYRCSRNIGDCQOTISRKGVKMGKCTIDCFACGSYENWIDNOKQF 409
Db 328 KTCNCRNYSRN---LYCGNGLDICOETIRVIGHNIVIGSECSKSYWCRRYKKWIDNOKKEF 384
```

```
QY 410 DKQKTYKEISDGGRRKRAVGTTK---YEGYEKSFYEKLKNDGYVDAPFLGNNEK 466
Db 385 LKQKKCKEENELMSKKQ-----SKYIVYEGYDKEFKILKSEVVGDLDFELINDER 440
QY 467 ACKDIT-DGGKINFEVNSGG-GVVGSGSGTSGASGTNDENKCTFYRSEYQCPQDCGV 524
Db 441 EQEFSNDLGKIDFETVDGGVGAIAG-----GASDINSNKN--TFSHSYCEECPCGCV 493
QY 525 QHKGNQNERKTYKKMWSLYKRPINGKMYLLKSLKYVDMMLKKMKFECLTQNSS 584
Db 494 E-LIGNEWKERKN-GECKGGRYNIPTKTHNVIPVLSFGDEHKETLEKIQFC----- 545
QY 585 DGSVSVYTTGASGNSSEKELDYDEMKCYKHNEQKVAVQGEVEDEDELKGAGLITP 644
Db 546 -----AESNDSKLTEDMKCY-----YDKREY-----VCTLE 574
QY 645 NPKRKEVSEAKSONNHADIOKTEHDFEYVYVAHMLDSIHMRTRKLSKCSIDSKTYKCR 704
Db 575 NRNKSEEDPE-----EIOKTFHNFEYFWIRHLNDSIEMBDK-INNCIEKAKEGCK 625
QY 705 NGCNKKCCCFEKWKYQKETEKKPIKDHFKTQEG--TPEGYFTTLELLKLOFLKEDTEE 762
Db 626 NECITDCGCFQRMWIGKKKEEWGEIKHKKTODGFSIFGNNDPVLBNVLTNDELFDOTE 685
QY 763 NTENS-----LDAEAEELKHLO-----KIL 783
Db 686 AYGNSQKIQIGIKDLAKKKTQAADATQOKNTIDLLFEYDSEAEKCKKIOEBQPKPT 745
QY 784 KLEN----- 787
Db 746 KYRNPYCGNNTYDALAGVAKLOQEAKEQLDRNDSRSALKANASQKYSNOGDPDFKK 805
QY 788 ----- 787
Db 806 NLCSITQKHSNAIGDSKPNCKNNKKEKERNVGEKKNGEYKMSHTDYLPRRQHFCTSN 865
QY 788 ----- 787
Db 866 LEHNTKSTGLTSDKALHSILGVDLLAAKKGEDIKTKLENDNRSSICRTMKKSPADIG 925
QY 788 -----ENNL----- 791
Db 926 DIIRGTDLMDINGDATGVQNLKIDFSKITEBELKQHPDKFNDKTYNDSKHTKLRSW 985
QY 792 ----- 791
Db 986 WEANRDQVWKAMTCPTKNGNIQCATPHDVIYIQRLRMWEAEMFCKEQSRLYEELLRD 1045
QY 792 -----AVNAGCTEQ 800
Db 1046 CGSGTTGKCNNDKCAKCKQCOQEYKTKIQIPWADQWNLISNKYQILYQAKIAALNGSTK 1105
QY 801 KTL---MDK----- 806
Db 1106 STTKKDDKKXVIDFLQKLEANYGTRPPEAHPDRPRRAATSKSDVYETTAGYTHQE 1165
QY 807 -----LNHELNDARKCD--CPLPEEDSKGRSADPSDI 840
Db 1166 ARTRECLGQNVFCNNNGNNEAFSLTPHEKHKACKENKRASSPE--LGRS-DSFDH 1221
QY 841 FIPREKEEDENDEDEVEDRD--ETAKETTES-----ATDTTSTL--DV 885
Db 1222 QTPPEDEVDHSSBEGEDESDEBEKEEVEYVDGADDEKAGAVSQPEASPTTKDQVKKP 1281
QY 886 CPVIGKVLTKDNESLDACSLKYG-GNNSRLGMRCV-----TP 922
Db 1282 CDIV-KELFSNVDPDLQACSKYKPGKN--YGNRCIPTKTSNDVTGDBGGSRVRYVSTP 1338
QY 923 SGEPTTSSDKNGALCVPPRRRLYIKIYVMATK-----TESQASGSEA----- 967
Db 1339 --ESGSNSDRNGATCIPRRRLRYVGLQEWANKHNTETVSOGAEATGASEAPAGGK 1396
```

QY	968	-SNTSGSTTPPDSKELALKAFVSSALETFFMLHNRKEE----	KAVAOEGACHOLPRAVE	1022
Db	1397	ESSSSGKEETPSDK---LRTAFIESAAVEFFFTFLDRKCKMELMAOKRAEALONG-GIDLVS	SSG	1452
QY	1033	EGSPYEPEDLKGKJTPDOFLRQMEYTLGADYDILJESGNTPTJASKDTPESSINDLNK		1082
Db	1453	DGDDP-NPQKLLNGVITPPRFLMLETTLADYDILYHGBN--TDSGNTGNSNN--N		1506
QY	1083	IVLLASGSTEOERKMKRYKEI--KNFRKCTERSAPNLVSHPOTWENNKGTYIWHGVC		1140
Db	1507	IVLEASGNKEDMQIQRKIBQILPRKNGPRLVFKSS--AQPRDKMNEHSAEISWKMIC		1563
QY	1141	ALT-----SKDKIANGVAKRQKILENBNLMD-----ANK-----	PKPQYQYTNV	1183
Db	1564	ALTYTEKRPDTSANGDE--NKIEKDEYVEKFFPGSTADKHOASTPTGYKYTOQYERK		1620
QY	1184	KLDSGSGSPPTTOGASONTPTLTHFRKPTJYRWMEEGESPCRRKRLKOIYD		1243
Db	1621	KLEDTSG-----AKTPASSDTP-LISDFYLRPIYRIYLEMGONCKERRKLQILKEE		1674
QY	1244	CKVENGDVGRCSGDGEACDSISTHDYSTVPSPCPG--CGKHCSSYRKRIEKKIYEFKQ		1301
Db	1675	CM--DGSDDKXSGGEGCCD--RRDJSNEVSADLBERSGSGNSGPRYKWKIKRREYXQ		1729
QY	1302	SNAGQOQKT---DATRANG--NTDKEFCYKLTETWPAKFLERLXNGPCKTKKEYGGDI		1357
Db	1730	ANAYASKOTKXYEEBSKSGAGINDHNKEFCVGLCTDAAFLRLKXNGPCKKDNENGGNDI		1789
QY	1358	DPEKSKTFOFTEVCGCPRKFKTNGONGN--CVSGLNCCGDKSIDAKELAKRSSTT		1415
Db	1790	NFGTTEELFRPAENCKCFKSSFFKNCRNGNCRSGDGDTEKENGNG--TITTGNEFTMGTOE		1848
QY	1416	DVWKRVSNDNTPEEGDGLDCAOAHNIFKGIKRDVWKGCVGADYECOTINERTDGK		1475
Db	1849	DVWVHVSDKANNEEGSGDGEACENNGIFLGIRKDKMKRYKCGHLICKOEGNGALINDQ		1908
QY	1476	EYIOIRALFKRWBNLEFDYKINDIKSIHCIRKGGSSKICNGCERNSKCLEKWIETKIAE		1535
Db	1909	QIITVIRALKRWMEYFLEDYKIKKRLKPCIEGNSGTCINCCKKNVGMGKIKRDE		1968
QY	1536	WENIKKRRNOYENKODPOYNVKS----JLEELIRLAVYNDQDVIKLCVFNESKGTL		1591
Db	1969	WTKIR--NHYLENNKGGDKNVTSLVTNVLETLYTIIAANDKRBOTSLDKTSLGNC		2025
QY	1592	ISNPN--KENDAIDCMLKKLYKAKNCG-----KPSGKQSDCKRPPLEPREED-Q		1642
Db	2026	PENSRKNDGNENDAIIDCMLKKLETKHECKTOGHENSVDOPHPRCGNGNP--PDEEJL		2084
QY	1643	NPEENTLEPPKFCPTTPOPEERKGGETCGKNEKKDEKKEESEEPAKESGPAAEAPAT		1702
Db	2085	LEENPVPWQPGFCPTPOQREPR--DDCKGLEKKRKEKQEPQPAEDGGA1----	VPS	2188
QY	1703	AESEETENPPEPFGTQPAAP--PSTPAP-PIYDTPPLRPOADEPFDSTLIQTTIPGY		1799
Db	2139	G-----PGSPEEDKGPVPRPAELPKPOEPDPSHAVIP--SLVTSILAMSV		2184
QY	1760	ALALGSIAPFLKKTKASVGNLFOLLIRPKSYDILPTLSSNRYIPIYVSDYKCKTYI		1819
Db	2185	GIGRAFFYFLKKTKISSVGNLFOLLIRPKSYDILPTLSPRRYIPIYSGYRCKRITYI		2244
QY	1820	MEGDSDEDKYAFMSDITDYVTSSESEYEELDINDIYVPGSPKXTLIEVLEP-----		1871
Db	2245	LEGDSGTDS-GYTHYSDIRTSSESEYEEMDINDIYVPGSPKXTLIEVLEPSSGNNTTAS		2303
QY	1872	-----SGNNTTASGKTPSDTRNDIONDGPSSKITTDEMNOLKEPISNMLQNOPDV		1935
Db	2304	GNNTTASGNNTTASGKTPSDTRNDIONDGPSSKITTDEMNOLKDEFTISQLOSEPYTE		2363
QY	1926	PUDYSGNSNTNTIITTSRHNVDNNTNTMSJDNNEENULLPSTIHDGULYGESEYSNV		1985
Db	2364	PR-----MLGYNVDNNTHTPTSHHNVEEKFPIMSHIDRNLUSGEYINDM		2408
QY	1986	-----NMVN-----SANDPIRINDNNVYSGIDLINDLSGGRPIDIYDEVLKR		2028

Db	2409	FNSGNPFINISDSTNSMDSLTSNNHSPYNDKADYUGIDLDNALSgnH-IDIYDEMfKR	2467
Qy	2029	KENELFEgE-NTKRTSFOVAKTNTSDPIHNOLEFgHKWLDRRHDMCEKAKNKEDILNKL	2087
Db	2468	KENELFEgRKHTKTKTNTNYNAKPARDPRTNOINLgHKWLDRRHDMCEKAKNHNBERKL	2527
Qy	2068	KEEWKKNINNSGRTKYSMDKPSHNHVLNDVSIQIDMDNPKTKNEITgMDNOQKSTSD	2147
Db	2528	KELW--KRETHSGDI--NSGIPSGNHVLNDVSIQIDMDNPKTKNETgMDNPPKSTSD	2583
Qy	2148	TILDDLEKYNRPYYDYDEDDIYIHVDVDEKSSMDIYUDHNHNVSNNDVPTKHHIEMN	2207
Db	2564	TILDDLEKYNRPYYDYDEDDIYIHVDVDEKSSMDIYUDHNHNVNNNDVPTKHHIEMN	2643
Qy	2208	IYNNKKEIPEEYPISDIWN	2228
Db	2644	IYNNKKEIPEEYPISDIWN	2664

RESULT	8		
ID	Q9U0G6	PRELIMINARY;	PRT; 2209 AA.
AC	Q9U0G6;		
DT	01-MAY-2000 (TREMBLrel. 13, Created)		
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)		
DT	01-MAY-2000 (TREMBLrel. 13, Last annotation update)		
DE	VAR. MALP2.56.		
GN	VAR.		
OS	Plasmodium falciparum.		
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
OX	NCBI_TaxID=5833;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=3D7;		
RA	Devlin K., Bowman S., Churcher C., Harris B., Harris I., Lawson D.		
RL	Quail M., Barrett B.;		
SR	Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AL035475; CAB62897.1; -.		
SO	SEQUENCE	2209 AA; 250779 MW; 7A349F6FE9F7CB CRCt4;	

Query Match	38.4%	Score 4648.5	DB 5	Length 2209
Best Local Similarity	44.0%	Pred. No. 2.2e-245		
Matches 1054	Conservative 318	Mismatches 664	Indels 357	Gaps 84
QY	6	GGGCGODEDDAKVHLDFEGGKVHDEV----	HGEAKNVVSELSKLSLASL	LGEGATFVTKSM 61
DB	4	GSTGYDDDDAKNNFDRIGGVYDEIMKKRDADAKKIKELKGLSTPASI	LGESAGCTDDPC	63
QY	62	QTESKYTELLEAN----	SKRNPC-KKDGKGVNDFRSVEKAGYDNKIMKSCNGN	-TCA 114
DB	64	QTESKYTELISGSGSVARGHPCGVNSGKGEVSRFSERKSYDEKI	IGCSNSEGACA	123
QY	115	PERRLILCKNNPPNNMNSNDGSAKHLLLAECMAAYEESISTKTHPKI	DSKYP-GSDPP	173
DB	124	PRRISLCKNNFOKIN-NYSRKAHMLLDVCLAAHNEQSITHLKO	DAEETPSGHT	182
QY	174	MCMTLARSFADIGDIIIRGHDLYLGNKKKKONGKETEREKLEOKLKEIF	KIHNDL----	K 229
DB	183	TCTALARSFADIGDIIIRGDLRRDK	-----GKKKLEHLTKITFKIHSDVSSGS	234
QY	230	DKEA-QKRNGBEDPFFYLRLEDWMTANETWYGAMTCSKE--	LDNSS-FRATCNT	-GQ 285
DB	235	NEALDIERNGKE-NYYLREDWMTANETWEALTCDDDKLAASI	FRATCSDDSK	293
QY	286	GP-SQTHNRCRCDKDKAGNAGPKAGDGVYTIPTFYEDVYPOLRMFEL	MAEDFCRRKKR	344
DB	294	GSFSQANDRCRCKDKKNTDQ-----	VPTFYEDVYPOLRMFELMAEDFCRRKKR	343
QY	345	KLENTLEKCRGDKSDERYICSRNGYDCQITISRKGVAMKGCCTDCFI	ACGSYENMIDN	404
DB	344	YNIYVTCYRKKDNNSEERYICSRNGDECKRTAKITLKVYGGCTDCFI	ACPYEYEMDN	403

Qy	405	ORQOPDQK-KYTHEI---SBOGGRKRAVGTTKYEGYERKPEYKIKANDYGTVAFLGL	461
Db	404	KKKEFLQKQEKYIINVINGTSSSSKRTTAAAG--SNVNGYERKPEYKIKEDGYNVNDLGL	462
Qy	462	LNNEKACKDITD---GKIKFKEVNSGGVVGSGSGTGAAGTNDENKGTFRSEYCOOP	518
Db	463	LNNEKACODIDKDEGGRKINFK-----DHGDINNKKDKGTFRYBKRYCOOP	508
Qy	519	CPDQGVQHK-----GQNMQRKTYKKMKWSLYLRPNKMYLLLSKLTVYDKMMLTKP	572
Db	509	CPYGVKKNNGNSGGGNKMKDEKHESDKTRIKLYPRSGOGGTPIKILSSEGEKEIKE	568
Qy	573	NMKRECU--TONSSSGSVSVTTGASGSGNEKKELDEKMKCKYKNHNVQKVVNOGEVEDD	631
Db	569	KIDDCITKYTN-----GTDSIDSSILDDPMKYEYIDELYEGOEGBDDVD	615
Qy	632	--DEL-KGAGGCLILPNPK--NKYEVSXSONNHADIQKTFHDFEYVWVNMMLKDSIH	685
Db	616	RYDELIVETGGGLTILKKEKKBOEKESDAKSONDDELQKTFHDFEYVWVNMMLKDSIY	675
Qy	686	WTRIRKLASCSIDGKTMKCRNGCNKKCDCEFEKWVKQKETEMLPKDHFKTQESIPGYVFT	745
Db	676	WTRKRLKLCXONGKKKCGKIKCNGDECECFQWYERKKTETWNIKHDFVQKQIPGECYFT	735
Qy	746	TLEILKLOPLKEPTENTENSIDAEAEELKILQILKYLENNNLVAVNA-----	797
Db	736	TLEGLVLOLEFENESADCKOXSIDAKTIOHLQIKILBEVKQKOEETAGCGGCVASD	795
Qy	798	TEQKTLMDKLNLHEINDATCKD---CPLEBEDSKSGRSADSPDRTFIPREPEKEDDENE	854
Db	796	NKKEITIMDKLIDYEKKIATBCIEKHKRPPRE-GAGRS--DTSRDSPPSRPAOEVGDSE	853
Qy	855	DODEDEVARDDE-----TAKETBESANDTTTSLD-	884
Db	854	DEDEDE--DOEEHPDDGKDGDANEEREAENHSDQEDKOTLDAVENTEYEGPGRPATPYD	911
Qy	885	-VCPITYGVLTIKDNESLQDACSILKYG--GNSRLGRWCYTPSEPT--SSDK-CAICVP	939
Db	912	KVCDIYKLTFS--GNDGDAKCGTRKYDKYGEKRPNNKCI--PSGDKAAPSDBNOSTCVP	968
Qy	940	PRRRRLYIKKIVDMATWTESPOASGEASSTGTPRPSKEALLKAFVESAIIETFLM	999
Db	969	PRRRRLYVGLTKWA--SGNTYVSGAOQTPQDGTSPDNK--LRDAFIQSAIETFLM	1024
Qy	1000	HRHYEEKKAAVQEBAGICLPRVEGSEYDPEKLAE-GKIPDGLRQMFYTLGDTROL	1058
Db	1025	HKYMKDEIEKKDOOKNGLV-ANTSNGKHQKLSQSGIIPEDFKROMFYTLGDTROL	1083
Qy	1059	FGSGNDTYSVKOPRSSNNLKIYVLASGTSEDBREKNNYKXKKNRCKSTSRAPN	1118
Db	1084	F-----GKDI--SGDKNMOTI-----EKKINGILP-KNCTSPRAKKNPT	1120
Qy	1119	LVSHQPTWMENNGKYIWHGVNCAITSKDKIAKVEKKPKOTENPENLMDANKKPKPOY	1178
Db	1121	-----EWSNGEHIINMAIICALITYDNTAS--DKPTONEKYKALMDQNNKPK-NDY	1172
Qy	1179	QYTNVNL-DEMSGTSPTTQTOASSDNTPTTLTHFVKRPTTYRMRVEDEGSECBRRKRL	1233
Db	1173	QYSSVTTGGEBAGEQLOSTDSKAAAREKPTPLDSFTKRPYFRUYLEBQWONCKERRKL	1233
Qy	1238	KQIYVDCKEVNGDVGRCGSGGECADSISTHDYTSVFNCPGCGKHCSSYRKMIERKIE	1297
Db	1233	KDICEYRGDENTIRYSGSGGECCKNNLPENBPTFDELYPICAUKYCRYYKKMYTKTE	1293
Qy	1298	FHKOSNAVGOAKTDATRNNGTNTDFEKCCTLETWPDAAFLERLKNGPCKTNKEYGDDI	1357
Db	1293	YEQDEKITYVOOKRQKDATSDNKNKYDSCMDGKLQYASIESFLEVLQ--CK--KONGECTI	1348
Qy	1358	DPEKDSKTPQHTYECGCGCPKFKTNCQNGNCVSGLNGNDGOKSIDAKEIAMRSSTDY	1417
Db	1349	KF--NGGQTFQHTEDCKSCSKSFRIKCNDDCKSGGNTKVKCDGKTPIDAKAEIAMINSPOEV	1407

QY	1418	VMISDSDTMTFEESDLDLKADCSOANIFKGIKRWKMGCGVGVCSIDE-----GQINMR	1471
Db	1408	TMLVSDGATGFGKDDDLKEACBESKGIETBEGIKRQMKCDNWCGVYCKPREGNRETVRBEK	1467
QY	1472	TDGKEYIQIALFERNWENTLEQYNNIKDSHISCIKGGSGKSCINGCEKNSKCEKMTIEK	1531
Db	1468	NDDKHITITRALYHWQNTLDDYKKIKKHISHICTDGDGSGTCQKCKCSGVEMLPK	1527
QY	1532	KIAEMENIKRPNQYEN--KQRPDYVKSILEELPKIAVVDQD---NVIKLCYFENSK	1587
Db	1528	KREEMOQIKRFLKQYKNDKLIDEDFNLRSCTLEFVLQIOAAVGEDEKFKVILSVFDSOC	1587
QY	1588	GCTLISNTQ--NKNENADICMKIKLIGVAKNC--PEKSGEK--QSDCKRPRLPREEDQN	1643
Db	1588	GCSAIASSQKKNGEYKDAITCMILKEEKNKCKEDHSGEGQETEKRCQSPSEVEDDTL	1647
QY	1644	PEEWTLPFPFC-----PPTQPEEKGGTGCNKKEEKDEKKEESEEPAKEESGPAEE	1698
Db	1648	HEBEVVAPELCKDYIKAP--EPREEK--AC---DPAITTKERS--PATSGKETME	1698
QY	1699	P-APTAEEETENFNRPPPGTGAAPSPRAPRPTDTPRLPRQADERP--STILQTTI	1755
Db	1699	PVTPQDQSPDKT---PKEKGPKPKPSLQPQPR-----QKRETPEDMKKVASASAF	1747
QY	1756	PFQVALLGSIAPFLFKKTKKASVGNLQIOLQPKSDYDIPFLKSKNRITIPVSRYRGK	1815
Db	1748	PMTGVAFMLISVYMKKKSKPRV-DLFVSMEIQDNDYMPFLKSKNRVYPSGKYRGK	1806
QY	1816	TYIWEEDSDDEDKAFMSDITDVT--SSSEPEELDINDIYVGSPPKYKLIENVLEPSCN	1874
Db	1807	TYLVEEDSGDGS--GYIDNHSDTISSESETEBEDINDIYRQSPKRYKTLIEVLEPESKR	1865
QY	1875	NTTASGKNRTPSDTRNDIQNDGIPSSK--ITDNEWQOLKEEFISNMLQNDPNDYNTSGN	1933
Db	1866	ETNS-----GQIPNDNTPANKRPITDEEWNTLKDEFISNMLQSTQNTPEP-----	1909
QY	1934	SSTNTNTTTSRKHVQVNNTN--TMSXDNMBEMLILPSHDGULYSGEEYVAVNM-----	1987
Db	1910	-----ILHDHMHNTHPNMLYEDKHEEKFPJMLIDRMILIREEYSYMSXNSGNN	1960
QY	1988	-----VNSMNDI--PFINRONVYSGDILINDSLSGSKRPIDYDEVYKRENEI,	2033
Db	1961	GSYGISPISDNPISDLSDKAGPTSGHNHLYSGDILINDLSC--DYDIYDEMLKRENEI,	2018
QY	2034	FGTENKRTSTONVAKATTSNDSPILHNOLELFHKYLDHRHRCRCKMKKKBEDILKLEENK	2093
Db	2019	FGTHHTKHTTNNRVAIQODDPIITQNLNLFHKYLDHRHRCRCKMEHNRDLKLEIWDN	2078
QY	2094	ENINNSGRKYTN-----SDNKRPSNH-----VINTDVSQIQDMD	2126
Db	2079	E--TSSN--:HSGNDIHPSDIHSGKLSDPSPDNNIHSDIPRYVLNSDVSQIQHMD	2136
QY	2127	NPKKRNETNMQNTQOKSTMTDITLDDLEKYNDRYYDYDFEDDIYHNDVSEKSMDDIYV	2166
Db	2137	NPKPTNEDNVDCN-----PVGNNIYV-----:IIII	2156
QY	2187	DHN-----NVTSNM--MDVPTKMHTEMINVANKKEIFEEXEYPIDINI	2228
Db	2159	DNNNQTFPSPNPBVENNTYVAPTNQIEMDVNNK--VVEKEPIPSMDI	2209
RESULT	9		
Q9NFB6	Q9NFB6	PRELIMINARY; PRT; 2163 AA.	
AC	Q9NFB6;		
DT	01-OCT-2000 (Tremblrel. 15, Created)		
DT	01-OCT-2000 (Tremblrel. 15, Last sequence update)		
DT	01-OCT-2000 (Tremblrel. 15, Last annotation update)		
DE	VAR, MALP4.01.		
GN	VAR.		
OS	Plasmodium falciparum.		
NC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
NCBI	TaxID=5833;		

OY	1817	YIIMEGSDP-EDKYAMSTTDVTTSSESEYEELINDIIVPGSPKYLIEVLEPSGN	1875
OY	1796	YIIMEDSSSGDEKFAYMSDPTDVTSSESEYEELINDIIVPAHPKKYLEVLEPSGN	1855
OY	1876	TASAGNTPSDPRNDONONGIPSSKTITDENMOLKEFISNNLON-QPNVDENDTSGNS	1934
Db	1856	TASAGNTPTSOTNDIPISTDTP-PLTTDEMNTLKDEFISQYLQSOPKADVPDYKSGSI	1914
OY	1935	SNTNTTSTRHNVNDNTWTWSRDMENELLPSIHDLNLSGESEYSYNVMVMSMDI	1994
Db	1915	PLNTQ-----PNTLYFNKPPEEKPFITSHDRDLXTGEGDISNHMSNTMTMO	1961
OY	1995	PINRDNNVYSGIDLINDSLSGCKPIDIYEVLYKRKENEFGLENTKRISTOVAKTTNSD	2054
Db	1962	PYYSVSNVSGIDLLINDALNG--DVDYDEILIKRENELLEFGTNHVQSYHSISVAKLTND	2019
OY	2055	PIHNOLEPHMKMLDRHRDCEKMKKNKEDILNKKEPMKENTINSKTYNSDNKPSHNHY	2114
Db	2020	PIHNOLELPHMKMLDRHRDCEKMKKNKEDILNKKEPMKENTINSKTYNSDNKPSHNHY	2073
OY	2115	LMTDVSIOQDMNDPKTKNEITNMDTNQDKSTMDTLTLDLEKENDPYDYFEEDDIYHDY	2174
Db	2074	LMTDVSIOQDMNDPKTKNEITNMDTNQDKSTMDTLTLDLEKENDPYDYFEEDDIYHDY	2121
OY	2175	DVEKSSMDIYDH--NNTSNNMDVPTKMHILEMNTVNKKKIFEEBEYISDTWNI	2228
Db	2122	N-----DHDASTVDSNNMDVPSRVQTEMD-VNTR-LVKEKYPIADVMI	2163
RESULT	10		
ID	097312	PRELIMINARY; PRT; 2169 AA.	
AC	097312;		
DT	01-MAY-1999 (TREMBLrel. 10, Created)		
DT	01-MAY-1999 (TREMBLrel. 10, Last sequence update)		
DT	01-JUN-2000 (TREMBLrel. 14, Last annotation update)		
DE	PEC1120C PROTEIN.		
GN	PEC1120C.		
OS	Plasmodium falciparum.		
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
OX	NCBI_TaxId=5833;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=3D7;		
RA	Oliver K., Bowman S., Harris D., Lawson D., Quail M., Barrett B.,		
RL	Submitted (FEB-1999) to the EMBL/Genbank/DBJ databases.		
DR	EMBL: AL034559; CAB39061.1; -		
SO	SEQUENCE 2169 AA; 246913 MW; 1D4BAECB7AFAE3BE CRC64;		
Query Match	35.9%; Score 4341.5; DB 5; Length 2169;		
Best Local Similarity	42.9%; Pred. No. 1.3e-228;		
Matches 1025; Conservative 294; Mismatches 656; Indels 413; Gaps	93%;		
OY	13	EDAKHVDFEFGKVHDEVHEGEAKNYVELKSGLASIIIGETAFYVSKMQESKYTELIE	72
Db	19	KQAKELDMIGQTVAHSVYHREDANRGKLYGLLT-----QAQFSNKERVH	63
OY	73	ANSKRNPPCKD-----GKGNDVDFSVKKEGAGYDNKCMKMSGMCAPRR	118
Db	64	IN---NPCLLDINYNTWNVSINVIPCCEKSVSRFEVSGGECDEKKIGKSGNCAAPRR	120
OY	119	LHLCKNFPPNNMNSDSRAKHDLAEVCGMAAKYEESESITKTHPKYDSKYPGSDPFMCYML	178
Db	121	LHVCGRNLDEQIRPH-TIATHTNLVDVCAQAFBEKSTISGYPRQYTKXDSGSTICYLV	179
OY	179	ARSPFDIGDIIRGDLVYGNNKKKONGKETEREKLQKLETFKKIHNL-----KD	230
Db	180	ARSPFDIGDIIRGDLVYGNNKKKONGKETEREKLQKLETFKKIHNL-----KD	235
OY	231	KEAQKRYNGDEDPRNYKLRBEDMWNTANRETVGAMTCSELDNSSFFRATC-NDTGOGSQ	289
Db	236	QALQARKYDDDGSGNTYQLREBWDMDNRKAVVAITICGAG-TSDKFRFKYCSGD---SD	290

QY	1261	CDLSISTHDYISVSPFNCPCGCRKSCSSYRWRMEKRTLEFKKOSNAAGCOQKT-D-ATRRNONT	1313
Db	1248	CSKLTVEKKDKIFKDLKEXPSCATPCPGYKIRKWRKRDYENKOKSAANEKTKTYENGNNCG	1307
QY	1320	FDEFCFTLETPWDPAAKFLERLKNKPC-KTNKEYGDDIDPEKDSKTPHTEYCGPCPKF	1378
Db	1308	GGNCVGTCLQ--ENADFLKRL--GFCSTTYENGVKRTD-DKTKRHKRHCDCPKF	1362
QY	1379	KTNCGNCGVSGLNCDCDKN--IDAKEIAKRSSTTDVYVARSNDNTNTEGDDJ	1434
Db	1363	NVNCNCNCSGGGKVECCNCRNNGTITIASDI-KNGDSFTEINMLVSDKPTTPEBGDL	1421
QY	1435	KDACAQHNRKIGIRKDYMKCGYCGVDIC--ECOTINENETDCKEYIQIRALFKRWENEL	1492
Db	1422	K-ACKNANIRKISRKDYMKVCVNICGYDVGCLKGGDNNGELDKOITILRALIKRMLLEYFL	1480
QY	1493	EDYKINDKTSHCIAKKGEGSKINGECKSKLEKIEKIAEMWENIKRFRNDYENNDQ	1552
Db	1481	EDYKIRKIKLKSCTEGESGSPCL-----KECDWITEKREKEMWTIKERYVDKTEMD	1534
QY	1553	PDYVWKSILEELIPKIAVVDQDNQNIKLCVFENSK--GCT--LISNTNNKN--DAIDCM	1607
Db	1535	GSNDLTFLFQ--GPFISLVEAKKVKYK--CKDEDEKIMGCTGTGTTHADKXENDEFITNL	1592
QY	1608	LKRLGVKAKKCPKPGSEKOSDC--KEPP--PLDEE-----DONPENTLEPPKFCPP	1657
Db	1593	ISKLOEKISECTOSSG---SDCTLSTENPSTLDEEPLVEDPNDAOKMIPKICGD	1649
QY	1658	TTQPEPKGGGTOGNKEEKKDEKESSEBPAAKEESPAABEPAPAESELETNPPERP	1717
Db	1650	VIREEKKEBEGGCPDLKKEEKKKKPEQ---TAEKPAABETKAE-----EKPA	1698
QY	1718	TGPAPAPSTPAPPTPDTPPLRQADEPDDSTLIQTTPFGALALGSIAFLFKKTKA	1777
Db	1699	QDEAGPP--PADPLP-----AAEPDPDITLQTTIPFGALALGSIAFLFKKTKS	1747
QY	1778	SVGNLFQIILPKSVDYDIPLKSSNRYIDVSDRYKKTYYIMEGSDSDKRYAFMSDTTD	1837
Db	1748	PV--DLFESVINIPKSDYDIPKLSPNRYIPYTGKRYGRKRYIYLEGDSGTDG--GYTDHYS	1805
QY	1838	VT--SSESEYEELDINDIYVSGPKYKTLLEVNLEP-----SGNNTTASGKN	1882
Db	1806	ITSSSEYEELDINDIYAPRAKRYTTLLEVLEBPGNNTTASGNNTTASGNNTTASGKN	1865
QY	1883	TPSDTRNDINDNGIPSSKITDNEMNOLKKEFISNMLO--OPNDVNDYTSGNSSTNTMIT	1941
Db	1866	TPSDTQNDINDOIPSKITDNEMNMLTKHDFISQYILOSQPDVNDYSSGDPPLANTO--	1923
QY	1942	TTSRHNVNNTNTTMSHDNNEENLLPSIHDGNYSGEYSYVNM--VNSMNDIPINRDN	2000
Db	1924	-----PMTLEFDKPEEKPFITSIHRRLTYGGEYSYVNMSTNSMDIPIHNSH	1972
QY	2001	NVYSGDILNDSLSGKRPIDIVYEVILAKKRENELEFGENTKRTSTONVAKTNSDPIHQL	2060
Db	1973	NIVYSGDILNDSLSN--VDIYDEVILAKKRENELEFGTNHAKRHTNTINAVKSSNSDPIHQL	2031
QY	2061	ELFKHMLDRHRDMCKEKKRNKEDILINKLEEMKNEINNSGKTYNSNDKPSHNVHLNTVDS	2120
Db	2032	DLFTWILDRHRDMCKEKKRNKEEVLNKLKEEM--ENETHSGNHTHSDS-----NKT	2055
QY	2121	IQIDMNPKTNETJNTMKTQDNQSTMDTILDLLEKYNDEYYVYEDDIYHADVVEKSS	2180
Db	2086	IQIDMNPKPIINQFTN-----MDTILEDELEKYNB--YYDVQDD--IYYDVH	2128
QY	2181	MDDIYVH--NNTYSNNNDVPTKMHLEMMINVNNKKEIFEEEPESIDINNI	2228
Db	2129	-----DHDSSTYDVSNNMDVPSKVOIEMD--VNTK--LVREKYPIDGVMDI	2169
RESULT	11		
ID	097324		
AC	097324	PRELIMINARY;	PRT: 2209 AA.
	097324		

01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-May-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE PFC00005W_MAL3P8.1 PROTEIN.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_TaxID=5633;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RA Bowman S., Lawson D., Basham D., Brown D., Chillingworth T.,
RA Churcher C.M., Craig A., Davies R.M., Devlin K., Felkett T.,
RA Gentles S., Gwilliam R., Hamlin N., Harris D., Holroyd S., Hornsby T.,
RA Horrocks P., Jagsels K., Jassal B., Quail M.A., Rajandream M.-A.,
RA Mungall K., Murphy L., Olliver K., Oyler J., Parkhill J., Paulsen O.,
RA Rutter S., Skellern J., Squares R., Squares S., Sutton J.E.,
RA Whitehead S., Woodward J.R., Newbold C., Barrell B.G.;
RT "The complete nucleotide sequence of chromosome 3 of Plasmodium
falciparum." ;
RL Nature 400:532-538(1999).
DR EMBL: AL034560; CAB39115.2; -
DR INTERPRO: IPR002048; -
DR PROSITE: PS00018; EF_HAND: UNKNOWN_1
SQ SEQUENCE 2209 AA: 250475 MM: 5F644E29B0CDAE98 CRC64;

Query Match 35.6%; Score 4302.5; DB 5; Length 2209;
Best Local Similarity 41.5%; Pred. No. 1.7e-226;
Matches 1008; Conservative 319; Mismatches 663; Indels 437; Gaps

```

OY      9 GTODEADAKRVLDDEFQKAVHDEVGEAKNYVELKGSLSLASTIGETAFTYVKSMTQ----- 63
       1 :| | | | | :| | | | | :| | | | | | | | | | | | | | | | | | | | | | |
DB     13 GIEDTAAKHIFRIDRIKIYHEKAKKAQEGRSOLGSKSLTKATFEKAPA---GGQTGCNTC   68

OY      64 ESKY---TELLEANKRRPPCKDKDGNDVDPRSVYEQAGYNKKKKCKSTGMTCAFRRLH    120
       1 :| | | | | :| | | | | :| | | | | | | | | | | | | | | | | | | | | | |
DB     69 ELKYWMHNVTYKGGKEYPCR--NGTE-KRPSVSGEGCDSDSKTSIGACAFPRFLN    124

OY     121 LCNKRFPMNNSSDSKAKHD-LLAECMAAKEGESIKTHYPKYDSKYISGFPCMTMLA    179
       1 :| | | | | :| | | | | :| | | | | | | | | | | | | | | | | | | | | | |
DB     125 LCVRLDENI--NNYRKINNDDTLADVCCLAAIHESGISGDHDKY--KE--NDSOLCTMLA    180

OY     180 RSFADIGDIIRGRDLGNKKKKKKONGKETEREKLQKLEIFKIKIDNI.KDKEA---OKR    236
       1 :| | | | | :| | | | | :| | | | | | | | | | | | | | | | | | | | | | |
DB     181 RSFADIGDIIRGRDLYRGN----NGK-----DKLEENIKTFGKTHEGI.KNGKTGIER    231

OY     237 YNGDEDPNFYKIREMWTANRETGVAGAMTCSELDNSSYFRATCDTG(G)SGSFTHNKRRC    296
       1 :| | | | | :| | | | | :| | | | | | | | | | | | | | | | | | | | | | |
DB     233 YG--NDPPEFOLRREDMWTANRETVKARICSPRADYFIKTVCSG---GKTPPGCGK    286

OY     297 DKDKGANAGKPRAGAGDVITYPTVEDDYPOYLRMPEEMAEDPCRRKKKIKILENLEKCRGK    356
       1 :| | | | | :| | | | | :| | | | | | | | | | | | | | | | | | | | | | |
DB     287 I-----DPS-VPTYDYIPQYLIRMPEEMAEDPCRLRKHI.LDNNAKKCKCEK    330

OY     357 DKSDEYRYCSRANGYDCEOITISRKGVRMGKGTDC(F)FAGSGSYEMIND(M)RRQFDKQ-KY    415
       1 :| | | | | :| | | | | :| | | | | | | | | | | | | | | | | | | | | | |
DB     331 HKDGKLLCYDLDFGPDQOTARGKNKKYKEHCIECYSCHDFHVHWID(M)KREFEKQKKY    390

OY     416 TKEIDSGGGRKKRAVGTTKYEG-----YEKSTYELKANQDYTYV(L)FLGLNNEKAC    468
       1 :| | | | | :| | | | | :| | | | | | | | | | | | | | | | | | | | | | |
DB     391 DEIKRAKGN---GTTTKETSNGRINNLVYGDEYSKLOOT-GVSVD(C)FLFKILDEAIC    445

OY     469 K---DITLOGKINFKEVNSGGVGSGSGTSGASTDENKNGFYRS.YCOPCPDGCVO    525
       1 :| | | | | :| | | | | :| | | | | | | | | | | | | | | | | | | | | | |
DB     446 KKHPEVEKTDVNF-----NENDDIFSHT(Y)GOACPICGLE    482

OY     536 HKGQMEWRKTKVKKMRMSKLYKPINGKRVLLKSLKLVKMDMLIKKNWKFCFLQNSSD    585
       1 :| | | | | :| | | | | :| | | | | | | | | | | | | | | | | | | | | | |
DB     483 ND-----SPMKP-----KREFCRCPOOI--RRJDD-----NESN    510

OY     586 -----GSVGSVYTTCGASGNSKEKELYDE-WKCYKANEQKV(V)GEVEEDDEL    634
       1 :| | | | | :| | | | | :| | | | | | | | | | | | | | | | | | | | | | |
DB     511 ETIKLVKRDGGQTNWEKLGGLGNGAKNNNIQETKWICY--YDKKKENSIIGSGDND---    564

```

0y	635	KGAGGLCLTPRPNKKNKEVSEAKSONNHAD	IQTETHPEFYVAHMHKJDSIHMRTRKSC	694
Db	555	-----YCLAKDKKNRFOLEIVS-----	-----FSLFNRWVTMLKJDSIDMR-KEIYKC	608
0y	695	ISDGKTYMKCRNGCKKKKDCFEKWNKOKETEMKD	IKDHFOTOGIEPGYV-FTTLEIILKL	753
Db	609	INNDGKSTCKNVCKKPCDCFOKMGVTRKTEBMQVKA	HYE-KEDFGKGLTPXYTLEWIDL	667
0y	754	QFLKEDPEENTENSLDAEHEALHKLQIKLAKENEN	NLVVNAAGTFOKTLMLKLNHLEIN	813
Db	668	STFPIKEAHKPE-----KPVOKMEIILKKQEN-	-----ISRWYKONNSITFKLOEILQ	716
0y	814	DATKC-----KDCEPLPEEDKSRG-----	RSAD-----PSPDIFIRPREKEDENEDDD	857
Db	717	EANNCLQKRDQCKRPQGSABEGVAKTGQRADEED	SDRPRDAGAEVDEDDADNDDE	776
0y	858	-----EDEVNRDE-----ETAKETTBGS-	-----ADDTTTS	882
Db	777	ITPRDLNIEVDDLDKSDPEDOVEBEKAKNDJDET	VEJAKETKEDDIDRKGEKQPEEVDN	836
0y	883	LDVCPYGVKLYLT-K-DNSLDDACSLKYG-GNSRL	DMRCVTP-----SGEPTSDSKNGAIC	937
Db	837	VKPCETIYKTLTTNDNDNTALNBACQKQYQYGEK	EPMMKCPPTTSGSEPTGSS---GSLC	893
0y	938	VPPRRRLYIKKIYDMA-----TKTESPOASGEA-	-----SSTSGS	973
Db	894	VPPRRRLYVPLTRLAGGGGTAAOSGSEBAQPV	IVTVO-POASGGNTQVAVSPGAASS	952
0y	974	TTTPDCKEALLKAVESAALETFFPLMHRKYEKK	AAQEGAGHGLPRVEEGSPETDPEDK	1033
Db	953	-----TSTESSOLLQAAIESAAVETFFLMHKKY	KVDKEKDEKERNMNVYVTSPPVN-DLYEK	1011
0y	1034	LKEGKIPDGFRLQRFYTLGVRDILFSGSNDT	SVSKOPSSNDLKNIVYLASGTE-1092	1022
Db	1012	LKKEIIPDPLRQMFYTLADYKOLTSG-----	-DKDKKNGYND-----ILSGDEL	1057
0y	1093	QEREKMNKYKEIKKFRKCSTERSAPNLVSHP-	-----OTWENNGKIYIHWGMCALTSK	1145
Db	1058	HERE-----KEIKG-----ALEKHFSNVK	PTPSGDADATWKENAPHIWEGICALTYK	1108
0y	1146	DKIAKGYEKKKQKIKENENLMD-----	ANKKRPQOYOTVNTYKLDENSGT-----	1191
Db	1109	ETMSGGGER-----QTEKSDVYKPFPGSTA	QTQGYTESYKTKTDVAKLENSGSGPRGN	1165
0y	1192	-SPRTTOASDDTPTTLTHFKRPRPYFWMPEME	SESCREKKRLKQIKVDCVENGD	1250
Db	1166	ESPMSKPPSSISEKKPTTDSFIKRPYRFLY	LEMEWETCREAKRLAQIKVDCGDENT	1225
0y	1251	VGRSGDGEACDLSITHDYSTVPSPNC	PGCGKHKCSSYRKWIERKIKTEPHKOSNAY-GOOK	1309
Db	1226	NRSDGGEFDEKVKYTKDVLDFEDNGSSCATC	SSSYRKWIERKIKIEEEDENAEYAEBOOK	1285
0y	1310	TDATRNNGNTPEDFECFTLETWPDAKFLERL	KNGCJTKNKEGGD-----IDPEK	1361
Db	1286	KNCVNGNNKGVNYCCKLE-ENAGFLQKL-GGCK	--KOSGEDNGNGNEDEXLANRQ	1339
0y	1362	DSKTFQHTETGCPCKFKTYNCQNGCVSG	LANGCDGKSIDAKETIAKRSSTTDVNRV	1421
Db	1340	PNVTFERAENCKPCSLIEIKCKNGVNGC	PTKECNGE-TVALEETKMNLDNGIMDLV	1398
0y	1422	SDNQTNFEEBDDLKADAOHNIKFGIRK	BDVWKGVCYGADICEDQNJNINRTQKREYIDR	1481
Db	1399	SDNCKNEIP-EDLKSCKDNINRTGRIEDW	CKGKFRDVCVCLNFNKHIDHKKNVILR	1457
0y	1482	ALFKRWENPLEDNKINDISHCIRKGE-	-SKCINGCEKSKSLEKWIIEKIMENI	1539
Db	1458	TLFRWLETFEEDYNRIOKTLKPCIE	NGKBEQCKGKCKGKEMNCDCVKKHVEKEKEMWIK	1517
0y	1540	KKRBNDOYENK-DOPDYNVKSIIEELPI	KTAIVVNDODNVIKLC---VFENSGCTLI	1594
Db	1518	KRRLEDYKNAAGSDDYKVKVSFLD--PQF-	-YNEVNAKVAVRPCDDLNAFERSIHNCNGPS	1573

QY	1595	TQNN--KENADIDCMLEKLVKAKKNCQK--PGEQSQSCCKPPLLPDEE----	ONPEN	1647
QY	1595	TQNN--KENADIDCMLEKLVKAKKNCQK--PGEQSQSCCKPPLLPDEE----	ONPEN	1647
Db	1574	SOKADVERRDVVYVCLLLEKLEKAKCKEOKHONSNGNQQCBGSTPPDDEEELLEEDQN	163333	
QY	1648	TL-----EPPKCPPTPTTPPEKGG-----ETCGNKEKKEKKEESEPAAKE	1690	
Db	1634	TVGKEKVGKNKPAALICGVDEQGEKKEEGEDQKAVTPDSDGEGKEDKESSEEEVSGS	169333	
QY	1631	ESGPAEPPATASSETETNFPEPEPGTGPAAFPSTPAFPPTDTPPLPAQADEFPDST-	1749	
Db	1694	DQSGPAPAPPEPSEPOEKAAPAPAEELLPGERPPK--PAAFPSTPEVPAQPL---P	SDNTS	1749
QY	1750	ILQTTPEGVALAGSIACLPFKTKTKASVGLPFIQIIPSDVDIPLKSSNRYIVY	1808	
Db	1750	DILKTTTPEGVALAGSIACLPFKTKTKST-DLNVIPPSDQGMPLKSPNRYIYT	1808	
QY	1809	SDRYKGTYYIMEGSDDEDEYFAFMSPTTDTV--SSESEYEELDINDIYVPGSPKTYLIEV	1867	
Db	1809	SGKRGKRYIYLEDSQDTS-GYTDHSDTSTSSSEYEEMDINDIYVPGSPKTYLIEV	1867	
QY	1868	VLEPSSGNNTTASGKNPSPDTRNDIIONDGIPSS-----KITDNEMNOLK	1910	
Db	1868	VLEPSSGNNTTASGKNPSPDTRNDIIONDGIPSDTRNDIIONDGIPSDTRNDIIONDGIPSS	1910	
QY	1911	KEPISNMLO--QRPDVNDYDTSGNSSTNTNTTSHRNDNNTMTMSDNNEEVLPLPS	1969	
Db	1928	DEFTSOYLQSQPDPVNDYDTSGNSSTNTNTTSHRNDNNTMTMSDNNEEVLPLPS	1974	
QY	1970	IHDNCLVSGEESYVNVAMVMSMDPIRDN-----NVSGIDLINDSL-SGGRKPIDYD	202333	
Db	1975	IHDNDVLTGEERYNTDMSNNNGIYPPSSNBRDLSGTVVPSGIDLINDSLNSGNOQIYD	2034	
QY	2024	EVLKRENELEFGEENTKRTSTQNVAAKTTNSDPIHNOLELHKWLDHRHDMCEKMAKKE	2083	
Db	2035	ELKRENELEFGEENTKRTSTQNVAAKTTNSDPIHNOLELHKWLDHRHDMCEKMAKKE	2094	
QY	2084	LAKLKEENKKNINSKRTYSNOKKSHNVLTVDYSIQIDMDNKRKTKEITNMDTNDK	2143	
Db	2095	LAKLKEEW--ENETHSNIHTSDS---MKTLTVDISIQIDMDKDKPKKEFN-----	2141	
QY	2144	STMDPTIDDEKYNDPYVYFGEYEDIIHYHVDYKESMDIYVDH--NNVTSNNMDVPTK	2201	
Db	2142	--MDTLEEDLKINER--YIVYODD--IYDVN-----DHDVSTVDSNAMDIPSK	2185	
QY	2202	MHEMNIVNNKKELEEEEPYDISDINI	2228	
Db	2186	VQIEMD-VNTR--LVKREKYITADVMDI	2209	
RESULT 12				
AC	P90580	PRELIMINARY;	PRT;	2647 AA.
AC	P90580;			
DT	01-MAY-1997 (Tremblrel. 03, Created)			
DT	01-MAY-1997 (Tremblrel. 03, Last sequence update)			
DT	01-OCT-2000 (Tremblrel. 15, Last annotation update)			
DE	FCR3-VART11-1 PROTEIN (FRAGMENT).			
GN	FCR3-VART11-1.			
OS	Plasmodium falciparum.			
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.			
OX	NCBI_TaxID=5833;			
RA	SEQUENCE FROM N.A.			
RP	STRAIN=FCR3;			
RX	MEDLINE=97154495; PubMed=9001213;			
RA	Hernandez-Rivas R., Mattei D., Sterkers Y., Peterson D.S.,			
RT	Wellens T.E., Scherf A.;			
RT	*Expressed var genes are found in Plasmodium falciparum subtelomeric			
RT	regions.*;			
RL	Mol. Cell. Biol. 17:604-611(1997).			
DR	EMBL; U67959; AAC47438.1; .			
DR	NON_TER 2647 2647			
QO	SEQUENCE 2647 AA; 303264 MW; 65FD700611348BC7 CRC64;			

Qy	1439	NDKISHC1KKEGSGC1NGCEKSKSCLKEK1IEK1K1EMENIKKRPRDYOKENQOPDYVK	1558
Db	1896	RKLLPFC1NNEKAK1GNG-----CVEQMINIKR1EWNNLKS-FNEQ-MDDTERBDR	1947
Qy	1559	-S1EELLP1AAVNDODNYKLCVFNSSGCL1SNTQNNK---ENPAIDCMLKLGK	1614
Db	1948	LRFPYDL1RQ1A1ITDKGN1NGVLKVLKYSKCACGNNSSONGKEGEBNDLVLC1LCKLEK	2007
Qy	1615	AKNCRKP--SGEKQSDCKEPP1LPEEEO--NPEENTLEPPKFCP-----	1656
Db	2008	AEKCDNENETSG1QOCCEVSPN1IEDEQPLEEENETVEHPRK1DDV1LKNHNOQR	2067
Qy	1657	-----PTQPREKCEGECGKKEEKQKDKKESEEPKKESSGAEEPA1ASETE1NF	1712
Db	2068	LKKNLVOP1LKRRKKR--KKRRKK1KKKQD1PHNLPCGAF1NTNP-----KTKP--	2118
Qy	1713	PEPOTGAEP1SPAP1PRTDTPPL1PQADEPFDST1QTT1PFGVALALG1AF1LK	1772
Db	2119	--PSSG-----KNWEH1PA1P-----ALVTS1LANSVG1GFA1FF1FLK	2158
Qy	1773	KKTKASVGN1E1Q1LQ1EKSDYD1P1LKSSNRY1PYVSDRYKGR1TY1MEGSDDEDKYAFM	1832
Db	2159	KKTKSTID1LLST--1PKSDYD1P1TK1SPKRY1PYSGYRGRKRY1LEGDG1TDS--GYT	2215
Qy	1833	SDT1DV1--SSSEKEELD1ND1YVPSGPK1TL1EV1LPP-----SGNNT1ASGN-1P1SD	1866
Db	2216	DHYSD1TSS1SSEKEEMD1ND1YVPSGPK1TL1EV1LPP1SGK1SGN11P1SGNNT1ASD	2275
Qy	1887	TRND1OND1PSSK1TQNNENQ1LKEF1SNMLQND1PND1PNDYTS1GNS1TNT1TTTSRH	1946
Db	2276	TQND1P1SD1P--P1TQDENN1TLKH1P1SMLQND1QKVD1PNDYTS1GNS1TNT1TTTSRD	2334
Qy	1947	NVDNNT1TMSRDMEE1NLL1P1S1HDGN1YSGEES1YNNWYNSMND1P1NRDNVYSG1	2006
Db	2335	NVDNNT1PTMSRHHV1DOK1P1T1SHRN1Y1TGEE1YNNWYNTMDD1P1NSHNNVYSG1	2394
Qy	2007	DL1ND1SLGK1P1D1YD1EVLK1RKEN1LFG1ENK1R1STQNVAK1TNSP1PHQ1LE1FHKW	2066
Db	2395	DL1ND1T1SGNH1D1Y1ELLK1RKEN1LFG1TNH1KQ1S1HSVAK1PT1RQD11HNQ1LE1FHKW	2454
Qy	2067	LDRHDMCEK1RKNED1L1NKLKEEMNK-----N1NN--SGKTYN--SDNK-----	2108
Db	2455	LD1SHDMCEQ1K1NDNER1KLAK1ELME1NETQ1CGD1NSG1P1SG1LSP1SD1NN1HSD1HPSD	2514
Qy	2109	-----PSHN-----HV1LNDV1S1Q1DM1DK1TNE1-----T1MD1TN	2140
Db	2515	1P1SGKQSD1P1SDNN1HSD1PYV1LNTD1VS1Q1HMD1NKP1NE1FTY1DS1NPQV1D1TYVDSN	2574
Qy	2141	QDK1STM1D1LDD1LEK1YND1PYV1YD1EED1--1YHVDV1EKS1MMD1YV1DHNNTV1SNNNDY	2198
Db	2575	PDNSSM1D1LDD1LEK1YNEP1Y-----DVQD1Y1NDV1-----DQ--ND1S1YV1D1NMDY	2620
Qy	2199	PTKMH1EMN1YNNKKE1FE1E1P1SD1WNT	2228
Db	2621	PSKYQ1EEMD1---NTE1FE1E1P1SD1WNT	2647
RESULT	13		
ID	Q25766	PRELIMINARY;	PRT; 2042 AA.
AC	Q25766;		
DT	01-NOV-1996 (TREMblrel. 01, Created)		
DT	01-NOV-1996 (TREMblrel. 01, Last sequence update)		
DT	01-JUN-2000 (TREMblrel. 14, Last annotation update)		
DE	3D1VARI1 (FRAGMENT).		
GN	3D1VARI1.		
OS	Plasmodium falciparum.		
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
OX	NCBI_TaxID=5633.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=96324414; PubMed=8670911;		
RA	Rubio J.P., Thompson J.K., Cowman A.F.;		

RT "The var genes of *Plasmodium falciparum* are located in the
RT subtelomeric region of most chromosomes."
RL EMBL J. 15:4069-4077(1996).
DR EMBL: U53324; AAB09769.1; -.
FT NON_TER 1 1
SQ SEQUENCE 2042 AA; 233669 MW; 89726CC777BB5C5A CRC64;

Query Match	33.7%	Score 4073	DB 5	Length 2042
Best Local Similarity	41.4%	Pred. No. 5.3e-214		
Matches 951	Conservative 253	Mismatches 581	Indels 470	Gaps 83

QY	139	HDLLAEVCMAAKYGBSEIKTHHYKDYKSPGSPFPCSTYLASPADIDGIRGRULYGN	198
Db	12	HNLLLDVCLAQYGGOSISONHGHOLSYPPDSPOSJCTELARASFDIDGIDYNGRULYGN	71
QY	199	KKKQONCKEPEKRELECKLEIKFRIKJHNDL---KDEAKQKRNGBEDDPFYFLJREDMMTA	255
Db	72	NR-----ENDLKEKLLKGYKKKTIYDNLVEKKKEAEATDYK-DDADPFYQJREDMMAL	122
QY	256	NRETVGAMTCSKSLDMSYSYFRATCNDTGGSPQJTNKCRCDKDKGANAKPKAGDGYT	315
Db	123	NRODYMAKITODA---HDSHYRKMGAD-GSITESAMRQR-----NVA	161
QY	316	IYPTFYEDYPOYLWMEFEAEADFCRKKKKKLEYLEQCSGKSKSDPYKCSNJDYCSOT	375
Db	162	DVPLNFEDVPOYLWMEFEAEADFCRKRKHKLKALOKCGOOGTCKDRYCDLNRDYCKRT	221
QY	376	ISRGKVMGAGCDDCFACSGSYENNIDNOKFOFK-KYTKEIS-----DDGQ	424
Db	222	ISAKHELVOGECKCKSVNCFIPGPMIDNOKOFEKQKKYTNELNKKHDEYTKEISNR	281
QY	425	RKRAVGQGTK-YEGYKSFYEXKLKNDGYGTADAFGLJLNNKACK---DITDGGKINF	479
Db	282	RKRSLSL---TFKNYGYDEEY-KIFDEYPDVDFLLDLKETACESOPUDEPRTISINF	338
QY	480	KEVNSGGVYVGGSGGSGASGTNDNKGTFYSEKQCSOPCGGYQHKGNGMEKTKYK	539
Db	339	K-----NYKNPDIFSHTEKQACAPMGM-----	361
QY	540	KMRMSKLYKPIKMGVLLLSLVVVDMMILKKMKKEPCLTQNSSDGSVGYVTGA---	596
Db	362	-----TMVLLMNVNQKTOGRNATYIKLFKFNKLIONTDPIPLPDTTKSNIV	408
QY	597	-----SGNSEKKELIYDMKCYTKHNEVOKVANVAGEVEDDELKAGAGLICL---LPNP	646
Db	409	EKYRNFCSNDDNNSDOINNMCOCH-----YDSKSKSGONDNCVEGTWQNF	453
QY	647	KKNREVSQAQSNNHADIOKTFHDFPYVVAHMLKDSIMPRKRLKSCISDQKT-----	700
Db	454	KKDOKV-----SYNAFWKAVSEMLDSSIMRAELDKLKNDKTKGKKNY	500
QY	701	MKC---RNGCNKKCDCEFKWVKOKETEMKPIKIDHKPTQEGIEGYEFTTLEILK-LQ	754
Db	501	KNCLNPLKLNGLNKK-----KKNRNGQ-KKPKRKK-----IDRKLICLLKHWLK	543
QY	755	FLKEDTEENTENSIDAEAE-ELKHIOKLIKLENNENMLAVNAAAGTQKTMDKLNLNHELN	813
Db	544	VLLNDVFLQ-----DMERKAGDPOHIAKIKELKKKXDEVNNL-SMNETIDFLLQDEEQ	597
QY	814	DA-----TKCDQCLPEDEKSRSRASDPDPIFIPPEKEKEDDEDDEDDEVRDE	865
Db	598	DAQCVSNNPKECKEYOKPPTDQAP-GAGPSPDICTDNLEIDSDQEBDDVSHUDEE	656
QY	866	ETAKETEGSA-----TDTTSLDVCPIVGVVLTKNDESLODACS	906
Db	657	EPEONPVEGSSSEEEKOEVUKDTAAVAPKODTOPKEVEVNPCKY-BELFKSTKNEPDAGCL	715
QY	907	KYGGNNRSLMGRCV---PTSGEPT-----SSDKGALICVPPRRRLATYK	948
Db	716	KYGN---YGMKCVCNHTSOKGSEPTLRGSHSVARSADGAPSDCKGALCIPRRRKLTLH	772
QY	949	KIYVMATKTESPOASGEASSTSGTTPPDPSKALILKAVESAAITETFFLMHRYKEEKA	1008

QY 287 PSOTHNKCRCDKDKGANAGRPKAGDGVITYPYEDVYPOFLRPFEMAEADPCRRKKKKL 346
Db 289 -----C-----GRNEGK-----VPTMLDVPYPOFLRPFEMAEADPCRRKRNLT 325
QY 347 ENLEKOCGRKDSDEYRCRSNGYDCBOTISRKGKVRMGKCTDCOFACGSEYEMIDNOR 406
Db 326 KSAKEKGGEBENT---KYCSLNGYDCTKVFEEKKDCSSDGCNCTACSNICLADYMLRQR 382
QY 407 KQFDKOR-KYTKELSDSGGRRRAVGGTTKYEG-----YEKSEYELKNDGYG 453
Db 383 NEFEKQKIKYTKYEIE-----KXSSSDKSNSNISMKNYYNEFEYENGKKEYE 428
QY 454 TYDVFGLLNNERKACKDITDGGKINFEKVNSGGVGGSGCTSGASTNDENKGTPIRS 513
Db 429 TLQNFPLNLNGMYCQ---KIEEBEV-----IDFNKDEDMVFHRS 466
QY 514 EYCOPCPDCGVQVHKGG-----IDFNKDEDMVFHRS 529
Db 467 EYCOPCPDCGVQVHKGG-----IDFNKDEDMVFHRS 526
QY 530 -----NOMERKTK 537
Db 527 EVFCSSTTNEGRNVQWKVCYNKNSDYNNCEMNISYKSTSDANVMLSVECFHSMAKML 586
QY 538 VKKRWMS-KLYKPIKGMVLLKSLKAYKD----- 566
Db 587 IDTKMEHOLKNCINTNTVTCES-KCICKECYEKWIKRHEHEKVKVNGNNRMSY 645
QY 567 -----MMIL-----KKNMKEF----- 577
Db 646 IYYNNLSRVDFSLFOVMFALDODEKGMDOFTEDLKKKPEPSTNPTGSDAIEFL 705
QY 578 -----CLTONSSDG-SVGSVYTTGASGNSK-----KELYDEMKCYAHNEVOK 620
Db 706 DHEKDNALTCBDNNSNESCDSVKVKTNPCKNPSASNNLVRYRLAEMMRVARKOLEK 765
QY 621 -----VNVQGE----- 626
Db 766 RGEINLKGDATKGTGYRGGPADGFKNVCSINQHNNTVQNNNRAVYOGPCTGKDGSGNG 825
QY 627 -----VEEDDELKAGGLCI----- 642
Db 826 VRMKIGTPMKGRIOIOMSAEDIYMPRRQHMCTSNLEYLQTKDGPLKOGDKLYVNSFLG 885
QY 643 -----LPNPKKKEV-----S 653
Db 886 DVLSSAKMDAGKIIELYKKQNNKSNLTPEDNESAARALRYSPADLGDIIIRGDLMDKNS 945
QY 654 EAKS-QNN-----HADIOK-----TFHDFY-YVWA-HMLKDSIHWRTKR 690
Db 946 DAKRLQTNLKEIFTKIKEELPEIDIKKYDKDGTDKLLREDMEANRHQV-----WRA-- 998
QY 691 LKCSISDGTWKCKHNG----- 706
Db 999 MKCAIENDKOKK-C-NGPIEDIYIPQRLRMTEAEMFECKEDSRLXNKLVLADCKCKKAK 1057
QY 707 -----CNKKCDCEFEKMWOKETEMKPIK----- 729
Db 1058 SCTOGDDCTKCAACDNYYNKKIKPWEBOEMKIKNKYALYKALDSVNGKEESKKKTAS 1117
QY 730 ----- 729
Db 1118 DAKDOOVVHFLAELIRKSGGKGNKVTYVSPPTPTNTLYSSAAGYIHHELGRVGCNT 1177
QY 730 ----- 729
Db 1178 QKECYSKNGYAKDPKYEAEACKDNRPKQAPAKKEDDADCVKPLDKDGETD 1237
QY 730 ----- 729
Db 1238 DIDCNOKYRAGDKYPGMDCNSQIHTTHNGCAMPRRQKLCVSGLRTDIRIKALEYIRT 1297
QY 730 -----DHKTOEG-----IPEGY---YFT----- 745

Db 1298 EFKSALETHFAMDRKEDNGEAEELKNGNIDEGFRQMTYTFGYRDIIEGRDISTH 1357
QY 746 -----TLELILQLOFLKEDTEENTENSL-----DAE 771
Db 1358 AYISGVSKYVTTILEKENDAKYAAKQNSNNELLDDWMDQHKDITWEGMLCATHTIKIDEE 1417
QY 772 EAEELKLOKILKL----- 785
Db 1418 KKEEIKKNYSYKKLINESPKGSNKVEDEFAKQPFLRPFLEMDEFCQAREEKAQVSCS 1477
QY 786 -----ENENLAVYNA-----GTOKTLMQ-KILNHE----- 811
Db 1478 DAKYDGCCKNTKSNVASCYACKVEYIYTKKVEYTKQKGFDEKITTDEKGYGFSKTD 1537
QY 812 -----LNDATCK-----DCPLPEE---DKSGRS 833
Db 1538 ASEYLLKKKCLDDTCNQKQVKNNTYWNTPKTYTNSNLEKRCQOPQPEPPPEBEGAR 1597
QY 834 ADPSPDIFTPRP-----EEREDEN----- 853
Db 1598 SDGSPRDTTPRPAAGSARSNTVSPPPRPAAGTVHEVAEVOEEDDEDGDLPEQDEBV 1657
QY 854 -----EDDEDDEVRD----- 863
Db 1658 EVAGAEEDLDVGVARILGRNTSPDEDEDAEEDDDDDQADPTTEVTGQEEETAE 1717
QY 864 -----DEETAKETTESGA-----TDTTSLDVCPIVGVLTJKNDESLODACLK 907
Db 1718 DHQOTTEYVDQEKAEEDKDGGETPOKETOPKAEVNPQDIY-KLTPTTETYLEACPTK 1776
QY 908 Y-GGNSNRILGRCVTPSGEPTTSSDKNGAICVPPRRRLYIKKIVDMATKTESPOAGSE 966
Db 1777 YVNGREKFPNMKCI-----SSGSDASGISICIPRRRKLTLHKI-----SGVD 1818
QY 967 AS-STSGSTPPDSKEALKRAFVSAAIEFFFLMHRKKEEKANAQOGAGHG---LPRVE 1022
Db 1819 TTVSSDEETTPITHDLREAFIQTAAVEFFELMHRKRIKKEKROBELONGFLFLPQAQ 1878
QY 1023 EGSPEDPE---DKLEKGIIPDGLROMFYTLGDRDILFSGSDJTTSV-SKDTPESSN 1077
Db 1879 KVPSPEDNPEHQKLLKSGKIPFEKFRQMFYTLGDRDILFSGSDJTTSV-SKDTPESSN 1077
QY 1078 DNLKNIVLLASGSTOEKREKMKYKEIKNFKKSTERSAP-----NLVSHQOTWENNGK 1132
Db 1939 KNIKDI-----SEKIKSVYEKSGEOTPPKPKGQTTTPEEMMOQNGE 1981
QY 1133 YIHWGMCALU-SKDKIAKGYEKKPOKIENTPENLMDANKKPKRPOYQYVNXKLDENSGT 1291
Db 1982 HIMNAMICALTHNTDTRQOVDDQVAGOLPENGN-----TPKNSOYQYKNV----- 2026
QY 1192 SPRTTQTOASSDNP---TTLJHFVKRPYFRMFEWGESFCREKRRKRLQIKVDKVEN 1248
Db 2027 -----TISVSSNGGPIGNIKIEQFASRPTFLRMLEWGECECRQYHKLIEKRECHKD 2082
QY 1249 GDVGRSGDGEACDSISTHDYSTVPFNCPCGCHGSSYRKWKIEKRIIEHKOSNAYGOQ 1308
Db 2083 N-RNCCDDDECEKEMPCPKDGSFETLKCPSACSKSYKMSIRKKDEFKQGAKEYQ 2140
QY 1309 KTDATRRNNGNTPDEKPECKTLETMPDAKFLERLKNGCCKTNKEVG-GDDIDFEEDSTFQ 1367
Db 2141 KKDAGEGNNDY--KEFSKTYLNNYNDAAFLNSLNGCPSKDDSDVODELKFDEKRTFG 2198
QY 1368 HTEYCGPCPKFKTNC-ONGNGCVSGLNGCNDGDSISIAKEIETAKRSSTTVVNVKVSNDT 1426
Db 2199 HETYCCKPSKTYCKCKENNNHODNKRPN-DCRNINISIADEIERSNSOTQVTVMSDSNT 2257
QY 1427 NTFEGDDLKDACQAHANIFKGIKRDVWKGYVGVYDLC--EQFINERTDKE-----YI 1478
Db 2258 NGNFYDLND-CIKAGIFKGIREDYVWKGEYCGVICTLEKTTNNEERVSASAKENDNKNQII 2316
QY 1479 QIRALFKRWENFLQDVNKNINDKISHCIIKKGEGSKCINGCEKNSKCLEKTIKKIAEMEN 1538

Db 2317 LIRVLEFKRWLESLFEDYKINIKINDISHCMKNDKSPCINGCQNKCNCEKVEKKEKSEMGK 2376
Qy 1539 IKRRFNDQYENKDDPD-YVAKSILEELIKRIAVNDQNVIKLCYFENKSGCTLLISNTON 1597
Db 2377 VBERYINQYRDNNSMEAFVKSFLLELPQIPVYTDKGHSDLTQKLLKLCNCSEKSEN 2436
Qy 1598 NKENDAIDCMLKKLKGKNCPCGKSGEQSDCKEPPRLPDEEDON-----EENPL 1649
Db 2437 SNEKDVYLCLLKLEDKAKNCQDQASGE-----PCPQTTSEPPDEDDILLFEENFV 2487
Qy 1650 EPPKFCPPPTQPPPEKSGFTCGNKKEKDEKKESEEPAKESGPAEAPAESEETE 1709
Db 2488 EAPNICPKVEEP-----EPVVEEKCDLAEAPKESSE-----TEENSGEG 2527
Qy 1710 TTFPEPPGIGPAPBPSTAPPTPPTPPPLP-----QADEPPD--STIIQTTIPFGVALA 1762
Db 2528 SNEQNPBRKPEEPPEPPPTPSSETDPBPAPPTIQPSQADQPNISIDLLISTPFGIALA 2587
Qy 1763 LGSIAFLFKKTKKASVGNLFQIOLIPKSDYDIPRLKSNRYIPYVSDRYKGTYYMEG 1822
Db 2588 LSIYFLFLKKTSSV-DLRLVNLIPKGEYGMPTLKSNRIIPYASDRYKGTYYMEG 2646
Qy 1823 DSEDEKYPASDPTDVTYSSSEYSEELINDIYVPGSPKYKTLLEVLEPSGNNTASGRN 1882
Db 2647 DSDSGHY--YEDTTPVTSESEYSEELINDIYVPGSPKYKTLLEVLEPSGNNTASGRN 2704
Qy 1883 TSDPFRNDIQNGISPKITTDENRQOLKEFTISNMION-QPNVDYPDYTSNSSTNTIT 1941
Db 2705 TSDPFRNDIQNGISPKITTDENRQOLKEFTISNMION-QPNVDYPDYTSNSSTNTIT 2764
Qy 1942 TTSRHNVDNNTMTMSRDMENLLPSTHDGNYLGESEYVNVNVMNDPIPIRDN 2001
Db 2765 TTSRHNVDNNTMTMSRDMENLLPSTHDGNYLGESEYVNVNVMNDPIPIRDN 2811
Qy 2002 VYSGIDLINDISLGGKRPIDYDEVLRKRENELEGTENTKRTSTONVAKTNSDPIHNOLE 2061
Db 2812 VYSGIDLINDALNG--DYDIDYEVLRKRENELEGTENTKRTSTONVAKTNSDPIHNOLE 2869
Qy 2062 LPHKULDRHRDCEKKNKEDILNKLKEEMKENINSKTYNSDKPSHNHVLNTDVS1 2121
Db 2870 LPHKULDRHRDCEKKNKEDILNKLKEEMKENINSKTYNSDKPSHNHVLNTDVS1 2923
Qy 2122 QIDMONPKTKNETIMNDTQDKSTMDTILDLLEKYNDPYDYDEDDIYHODVYKSSM 2181
Db 2924 QIDMONPKTKNETIMNDTQDKSTMDTILDLLEKYNDPYDYDEDDIYHODVYKSSM 2965
Qy 2182 DDIYVDH--NNTYSNMNDVPTKMHLEMTNVNKKELFESEYPTSDIWN1 2228
Db 2966 -----DHDTSTVDSNMNDVPSKYQIEMD-VNTR--LVKKEKTPRDVYMDI 3006

RESULT 15
025733 PRELIMINARY: PRT: 2924 AA.
AC 025733:
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
DE PFEMP1 VARIANT 1 OF STRAIN MC.
GN MCVAR-1 PFEMP1.
OS Plasmodium falci-parum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OK NCBI_TaxID=5833;
RN 11)
RP SEQUENCE FROM N.A.
RC STRAIN=MALAYAN CAMP (MCJ32 K+C+R+);
RX MEDLINE=95330812: PubMed=7541722;
RA Baruch D.I., Pastosloske R.L., Singh H.B., Bi X., Ma X.C., Feldman M.,
Tarasch T.F., Howard R.J.;
*Cloning the P. falci-parum gene encoding pfEMP1, a malarial variant
RT antigen and adherence receptor on the surface of parasitized human
erythrocytes.*;
RL Cell 82:77-87(1995).

DR EMBL: U27338; AAB60251.1; -.
DR INTERPRO: IPR000345; -.
DR PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_1.
FT PROSITE 104 104 S -> T.
FT VARIANT 105 105 E -> K.
FT VARIANT 139 139 D -> G.
FT VARIANT 362 362 K -> E.
FT VARIANT 503 503 S -> N.
FT VARIANT 581 581 S -> T.
FT VARIANT 593 593 M -> V.
FT VARIANT 617 617 I -> K.
FT VARIANT 631 631 V -> G.
FT VARIANT 658 658 D -> E.
FT VARIANT 722 722 K -> N.
SQ SEQUENCE 2924 AA; 335862 MW; 7077553BDAC82B26 CRC4;

Query Match 32.5%; Score 3937; DB 5; Length 2924;
Best Local Similarity 33.3%; Pred. No. 2,2e-206;
Matches 1036; Conservative 293; Mismatches 714; Indels 1066; Gaps 103;

Qy 1 MATSGSGGTODEDAKHVLEFGQKYHDE-VHGEAKNVYSELGSLSL SILGETAFYVK 59
Db 1 MGGGNGGGGTCKDKDAKHALDRIGEYVKKVENDAEKFKALKGNLQZKIGIGELASSPN 60
Qy 60 SMO-TESKTELLEANSKRNPK-----KDGKNDVDRFSYKEDAG;DNKKMKCSNGM 111
Db 61 PCKLVEDYVNNRKL--RKRYPCANRQTVRFSDEYGGQCTFNRRKSENIDN-----SIG- 112
Qy 112 TCAPFRRLHLCKNKNFPNMNSNDSKAKHDLAEVCAAKYEGESITKH;PKYDSKYPGSD 171
Db 113 ACAPYRRLHLCDYNELEMKR--TSTTKHDLIDYCAAKYEBGDSIKTH;TKHELNRPDK 170
Qy 172 FPKCTMLASFPADIGDIINGRDLYLGNKKRKQNGKETEERKELEQKLE;FKKIHDNLKDK 231
Db 171 SOLCTILASFPADIGDIIVAGKDLGLGYDKEKD---ERKLENNLIE;FKKIHEHLGTO 226
Qy 232 EAQKRYNGEDENFYFLREDMTANRETYWAGMTCSKELDNSSYFATY;INDQGGSGQSH 291
Db 227 DAKDHKKDEE-NYQLRBDMWTANSTYWKATIHAG-ESQKFPKTY--CSGE--WTD 280
Qy 292 NRCRCQDKGANAAGKPADGDVTVPTVEYDVPQVLRPFEEAEDPFC;IKKKKKLENLBK 351
Db 281 DKCRCKDEEGKN-----ETNEYPTTFDYVPQVLRPFEEAEDPFC;IKKKKKLENLBK 331
Qy 352 QCGKDKSDERYCSRNGYDCBQJTSRKGVKMGKCTDPCFACSGYE;WIDNRKQDFK 411
Db 332 NCRG-EKGNR-RYCDLNGVNCETARGAELFYAGDDCHRCVACDRFY;WIDNRKREFDK 389
Qy 412 Q-KKYTEISDGGGKRAVAGTTKEG-----YKSFYEKLKNDG;GYDAPAFGLLN 463
Db 390 OKKKYDEEIN-----KTHGTTITTGNGKINNLVGHFYKLLK-Y;PYDQTSLOKLN 440
Qy 464 NEKACKDITDGGKINKEVNSGGVYVGGSGGSGASGTNDENKGFY;SEYCOQPCPDG 523
Db 441 DEAIACK-----KPPN-----VGNKAST--VDNNEVNTTFS;ITTCCEACPMG 482
Qy 524 VO-HKGNQMERKTVKKNRWSKLYPRINGKMYLLKLSLVYKDMIL;KNNKEPCLTON 582
Db 483 AQKEKNGGMKAKKESCAKKERIFKKNSTDI---KILPREGRSKT;LEKLTQPC---- 535
Qy 583 SSDGSVSVYTTGASGSEKKELEIDEMKCYKHNVEQAKNVQKVVQGEVEDDELKGAGLCI 642
Db 536 -KDG-----QKIKNDIMWCHDD-----NCTDQTD-----SNDV 566
Qy 643 LRP-----PKKNVESEAKSQNNHADIQTFHDFEYVVAHMLKDSIHMRTKRLKSCISDG 698
Db 567 LGDMGNLTKEDKIMS-----YNAPFMMVMDMLDLSIKMWDEGR-CTIND 611
Qy 699 KTKKCRNGCNKQDCPEKWKVQKETEEMKPKIDHFKFYQESIPRGY-YFTIYLEILKQFLK 757
Db 612 KGTCTKCGNKKICQKKNVEQKKTTEWGIKIDHFRQKXIDIPKDWITDDELQTLMLKMDLL 671

Db 2772 ELLDKLKEEMNKDNNSGNINPSGNTPTSDIPSGKOSDIPSDNNIHSDIPYVLNFDVSIQ 2831
Qy 2123 IIMDNPKTKNETITNMDTNODKSTMDTILDDLEK-YNDPYYPDYEEDITYHDVDEKSSM 2181
Db 2832 IIMDNPKPINEFSNMDTYPNSSMDTILEDLDPENEPYDYDQDD-LYDVH----- 2883
Qy 2182 DDIYVDH--NNVTSNNMDVPTKMHIEMNIVNNKKEIFEEEPISDIWNT 2228
Db 2884 -----DHDTSTVDTNAMDEPSKVQIEMD-VNTR--LVREKYPIDLMDI 2924

Search completed: May 1, 2001, 13:08:52
Job time: 340 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2001, 13:03:32 ; Search time 29.5 Seconds

(without alignments) updates/sec
2367.160 Million cell

Title: US-09-508-967-1

Perfect score: 12100
Sequence: 1 MATSGSGCGTQDDAKHVLD.....VNKKKEFEFEYPISDIMNI 2228

Scoring table: BIOSUM62
Gap 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	418	3.5	1435	1	EBAL_PLAFC
2	315	2.6	1070	1	PVDR_PLAFC
3	301	2.5	2869	1	RBP1_PLAFC
4	277.5	2.3	1658	1	YH67_YEAST
5	269	2.2	1630	1	MSPI_PLAFC
6	265.5	2.2	1639	1	MSPI_PLAFC
7	265	2.2	1701	1	MSPI_PLAFC
8	264.5	2.2	1726	1	MSPI_PLAFC
9	262.5	2.2	1726	1	MSPI_PLAFC
10	262.5	2.2	3969	1	HRX_HUMAN
11	258	2.1	1701	1	MSPI_PLAFC
12	251.5	2.1	2476	1	ATRX_MOUSE
13	249.5	2.1	2349	1	TPR_HUMAN
14	249	2.1	1153	1	PVDB_PLAFC
15	245	2.0	2375	1	ATRX_HUMAN
16	244.5	2.0	2004	1	MOZ_HUMAN
17	238.5	2.0	1637	1	MRSP_STAU
18	238.5	2.0	1744	1	TANA_XENLA
19	238	2.0	1073	1	PVDA_XENLA
20	237	2.0	1466	1	SPA2_YEAST
21	233.5	1.9	1391	1	MSR2_DROH
22	232.5	1.9	2464	1	MAPB_MOUSE
23	226.5	1.9	1183	1	CNA_STAU
24	226.5	1.9	1875	1	MLP1_YEAST
25	222.5	1.8	1337	1	DEXT_STRDO
26	221.5	1.8	1682	1	MSPI_PLAFC
27	220.5	1.8	1070	1	PVDB_PLAFC
28	218.5	1.8	3924	1	ANK2_HUMAN
29	217	1.8	2843	1	APC_HUMAN
30	216.5	1.8	1679	1	Y109_YEAST
31	216	1.8	2748	1	Y109_YEAST
32	215.5	1.8	1957	1	YD66_SCHPO
33	213.5	1.8	1018	1	FNBA_STAU

34	213.5	1.8	1233	1	YF16_YEAST	P43597 saccharomyc
35	212	1.8	1370	1	2261_HUMAN	O14202 homo sapien
36	207.5	1.7	1616	1	P200_MYCGE	O49429 mycoplasma
37	207	1.7	1500	1	SPF5_STRGN	P16952 streptococc
38	205.5	1.7	1700	1	BAR3_CHITE	O03376 chironomus
39	205	1.7	1790	1	USO1_YEAST	P25386 saccharomyc
40	204.5	1.7	1139	1	HMW1_MYCGE	O49413 mycoplasma
41	204.5	1.7	1251	1	RBP2_PLAFC	O00799 plasmodium
42	203.5	1.7	2334	1	WAPA_BACST	O07833 bacillus su
43	203.5	1.7	2845	1	APC_MOUSE	O61315 mus musculu
44	202	1.7	1781	1	AKAC_HUMAN	O02952 homo sapien
45	199	1.6	1460	1	N159_YEAST	P40477 saccharomyc

ALIGNMENTS

RESULT	1	EBAL_PLAFC	STANDARD:	PRT:	1435 AA.
ID	EBAL_PLAFC				
AC	P19214;				
DT	01-NOV-1990 (Rel. 16, Created)				
DR	01-FEB-1996 (Rel. 33, Last sequence update)				
DT	01-FEB-1996 (Rel. 33, Last annotation update)				
DE	ERYTHROCYTE-BINDING ANTIGEN ERA-175.				
OS	Plasmodium falciparum (isolate Camp / Malaysia).				
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.				
OX	NCBI_TaxID=5835;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=90377299; PubMed=2204835;				
RA	Sim B.K.L.;				
RT	"Sequence conservation of a functional domain of erythrocyte binding				
RL	antigen 175 in Plasmodium falciparum."				
Mol. Biochem. Parasitol.	41:293-296(1990).				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation-				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@sib-sib.ch).				
CC	-----				
DR	EMBL; X52524; CAA36756.1; -.				
DR	PIR; S11561; S11561.				
KW	Antigen.				
FT	DOMAIN 159 1104				
FT	ESSENTIAL FOR BINDING TC				
FT	ERYTHROCYTES.				
FT	VARIANT 1031 1031				
FT	E -> V (IN STRAINS FCR-3 AND ITG).				
FT	SEQUENCE 1435 AA; 167389 MW; 32A4309021BIC3D6 CR; 64;				

Query Match 3.5%; Score 418; DB 1; Length 1435;
Best Local Similarity 17.4%; Pred No. 3.3e-11;
Matches 337; Conservative 259; Mismatches 555; Incls 782; Gaps 85;

QY	392	FRACGSEYEMDINORQKQKRYKEISDGGKRAVGGTGYEGYEKSEYK	---L 447
DB	17	YFAKANEDYKNEKEFLDYKKEKFEELD	-----KKKYGNGQTKDKI:FTFLENKLDIL 70
QY	448	KND-----GYGT---VPAFLGLL---NNEKACADITDGKINKREYNSGGGVGGG	--- 492
DB	71	NSKFNKRKRSYGTDPNDIKNLSIKKHNEEFNN	-----NFOSLSTSLIKONKY 123
QY	493	-----SGTSGASGTNDENKGFYREYCOPCDCGVOKHGNORERKTKVKKRMS	544
DB	124	VYINAVRYSRIISFLDSRIINNGRNTSSNNEVLSNCR	-----EKRKGKIDCKKKNDRSRTV 179
QY	545	KLYKPIGRMIVLLKSLKVK	-----DMILKKN---WKEFCLQON 582
DB	180	CI---PDRRIQCIYVLSIKITTYTKETMKDHFLEASKKESQLLKKNNIKYNSKFCNDLK	236

QY	583	SSDSVCSVVTGAS--GSGNSEKLEYDENKCXKHNEDVCQVYNGVEEDELGAGGL	640
Db	237	NSFLDYGHLMANGMDTGGISTAE-----NKIOEV-----FKGAHG-	273
QY	641	CLPNNPKNKEVSEAKSONNHADIOTKTFHDFYVVAHMLKDSIH-----	685
Db	274	-----EISHKIKNFRKEMWNEFRKMLAEALSEHKNNNNKINPDELOITQW	323
QY	666	---W-----RTRLKSCISDCKYMKCRN-----GCKNKC--DC--FEKWQKET	723
Db	324	IKEMHGEFLERDRSRKLPRS-----KCKNNTLYEACEKECIDPCPKYKDWIIRSEF	375
QY	724	EMKPIKHFQKOEPI--EGYFTTLELILQGLKDTTEEN--TENSLEAEAE--EL	776
Db	376	EMHTLSLEYEOK--VPEKNAENT-----LIKISENKDAKVSLLLNCCDAEYSKYDCD	427
QY	428	KHTTTLVK-----SVLNG--NDNTIKERHIIDDDSKGCG-----	462
QY	835	DPSDFIIPPEKEDEDEDDEDEYRDEDETAKEETEGSATDTTSLDVCPIGVKYL	894
Db	463	-----DKN-----SVDTNKTV-----	473
QY	895	KDNESLDADCSLKTGGNNSLGRWCYTPSGEPTSSDKNALCVPRRR-----RLY	946
Db	474	-----WEC--KNPIILSTKO--VCVPRROELCLGINDITY	505
QY	947	IKKIIVDATKTESPOAGSEASSSTSGSTPPDSKHALKAUFESAI--ETFFLMRYKEE	1005
Db	506	DKNLL-----MIKHIILAIITYSRLLAKKRYKN	534
QY	1006	KKAAVAGEGAGHLPRVEBSEPEYDPEDKLKEGIPDGLRQMEFYTLGDRILFSGSNDT	1065
Db	535	-----DDK-----EVOCKINTKFPADIRDIIL--GGRDY	559
QY	1066	TSVSKDTPSSSNDNLKATIVLLAGSGTEOEKKNKKEIKENRKRKSTENSAPNLVSHPT	1125
Db	560	W-----NDLSNRKLVGRKINRNSKVVHRNK--KNDRLEFR-----DE	592
QY	1126	WMENNGYIWMHGMVCAITSLKSDKILAKGVEKKPOKIEKPNMLMBEANKKPRPQLOYTNNKL	1185
Db	593	WMKATIKKADVNN--VISWVFEDKTV--CKEEDIEH-----	622
QY	1186	DENSGTSPRTTOASSDNTPTTLTHEVKKPTYFRFMEEGESFCGERKKRLKQIYVCK	1245
Db	623	-----IPQFRFMSMGDDYQCDKMKMETLKEVCK	653
QY	1246	VENDVGRCSGDGACDSTIDHYSTVPSFNCPGCKHCSSTRKMYERKKIEFHKSNAV	1305
Db	654	EK-----PCEBD-----NCKSKCSNKEWISKKEEYNKQAOAY	687
QY	1306	GOOKTDATBRNNGNTFDEKFCFKTLETMPDAKFLERLKNCPCTNNKEYGGDDIDFEKSDKT	1365
Db	688	OEYO-----KGNNTKMYSEF--KSIKPEYLLKKTISEKSN-----LNFDEPKE	729
QY	1366	FOHTEY--GCPGPKFK-----TNCQNGNGVGLNGN-----CDGKSIDA	1404
Db	730	ELHSDYKNNKCMCEVADVPISIIIRNNBOISQAVPEENTETLAHRTMPSISEGKGBQ	789
QY	1405	KEIAKMRSSTTDVVMRVSDMTTFEGBDILKDACOHANIFKQIRDV--WKGVYGVADI	1462
Db	790	KE-----RDDDSLKISIVSPENSREPETAQD--SNLLK--LKGDVDISMPKAVIGSSP	839
QY	1463	CEQININERDQKEYIQRALFRMWENFLBEDNKNINDIISMCIKKGBESKINCENKNS	1522
Db	840	NDNINIVTEQGN-----ISGV--NS	857
QY	1523	KCLEKMTIEKKIAEENIKKRFNDQYENKQDOPYNKSIIEELIPKIAVYNDODNYIKLCV	1582
Db	858	KPLSDVDRPP-----KRLLEDQ--NSDESEETVYN--HISPSPTNMD-----	897
QY	1583	FENSGKGLTISNTQNNKENDAI-----CMILKLGVAKAKCPCPSGKOSDCKEPP	1635

Dd	898	-DSGSGSATVSSSSSSSNGSLSDDDRNGDTPVRQDTANTDEYIVKENADKDEDEKG----	953
Qy	1636	LPEDEDONPEENTLEPPKFCPTTOPPEEK-----GGETCGNKEEKKDKKEESEEP-	1688
Dd	954	-ADEERHSTSE-----SLSPEEKMLTDNEGGSNLHEEVKHEHTSNDVNQSG	1001
Qy	1689	-----KEESPAADEPAPTAESETEENFPEPPTGTGAAPTSPAPPPTPIPLRPQ	1741
Dd	1002	GIVNMNVKEKELDLOTLENSSSIDECKAHEBELSEPMLSODMSNTPG-----	1048
Qy	1742	ADEPFDSTLOTTIPFGVALAGSLAFLEKLKKTASVGNLFQIILQIPKSYDIPTLKS	1801
Dd	1049	---PLDNTSEETTERI-----SN	1063
Qy	1802	NRIYIPVSDRYRGKYTIWEGSDDEDKAFMSDTTVDVTSSSESEYEELDINDIYVGSPRY	1861
Dd	1064	NEX--KVNR-----EDERTLTREYEDIYAKSMNESDGCXY-----	1100
Qy	1862	KYLIEVLVEPSGNNTTAGSKNTPSOTRNDIQNDGIPSSKITDENYNQLKEPFISMLNQ	1921
Dd	1101	-----DENGLSTVDESEDA--EAAMKGN-----TSEMSHS	1132
Qy	1922	PNDVNDTSCGSSNTINTTTTSKRNVONNTYTMSROMMEENLLPSTHDGNLYSGERY	1981
Dd	1133	SCHIESDOOKNKMKYTGDLGT--HYONEISVPATGEIDEX--LRSEKSKIHAEE	1186
Qy	1982	SYNVNMVNSMNDIPINRDDNNVYSIGIDLINDLSGCKPIDIYDEVALKRKENELETCTENTR	2041
Dd	1187	RLSHDDIKHN--PEDRSNLTJHLKDIRE-----ENERH	1219
Qy	2042	TSTONV-----AKTNSDPINOLEFHFKMDLRHRDCMKMNKEDILNK	2086
Dd	1220	LFNQMINISOERDLQKHGFHTWMNLHGCVSERSQINH---SHHCNRDDRGSGNSVTLNM	1276
Qy	2087	LKEENKKEINNCSGTYNSDNKPISHNHVLNTDVYSIQIDMDNKTKTNELTMOTNDKSM	2146
Dd	1277	RS--NNNFNPISPRMYLDX-----KLDLDIYENRNDSTYTKELIRKLAEI	1320
Qy	2147	DTILDLE-KYND	2158
Dd	1321	MCENEISVKYCD	1333
RESULT 2			
PDDR_PLAYS			
ID	PDDR_PLAYS	STANDARD:	PRT: 1070 AA.
AC	P22290:		
DT	01-AUG-1991 (Rel. 19, Created)		
DT	01-FEB-1996 (Rel. 33, Last sequence update)		
DT	30-MAY-2000 (Rel. 39, Last annotation update)		
DE	DUFFY RECEPTOR PRECURSOR (ERYTHROCYTE BLINDING PROTEIN).		
GN	PVDR.		
OS	Plasmodium vivax (strain Salvador I).		
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
OX	NCBI_TaxID=5856;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=91187056; Pubmed=1849231;		
RA	Fang X., Kaslow D.C., Adams J.H., Miller L.H.;		
RT	"Cloning of the Plasmodium vivax Duffy receptor."		
RL	Mol. Biochem. Parasitol. 44:125-132(1991).		
CC	-1- FUNCTION: BINDS TO THE HUMAN ERYTHROCYTES DUFFY BLOOD GROUP DETERMINANT.		
CC	-1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.		
CC	-1- SIMILARITY: HIGH, TO P.KNOWLEDSI DUFFY RECEPTORS.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/		

or send an email to license@lsb-sib.ch.

CC -----
 DR EMBL: M61095; AA63423.1; -
 KM Malaria; Receptor; Glycoprotein; Signal; Transmembrane.
 FT SIGNAL 1 20
 FT CHAIN 21 1070
 FT DOMAIN 21 1007
 FT TRANSMEM 1008 1025
 FT DOMAIN 1026 1070
 FT CARBOHYD 183 183
 FT CARBOHYD 255 255
 FT CARBOHYD 351 351
 FT CARBOHYD 420 420
 FT CARBOHYD 715 715
 FT CARBOHYD 787 787
 FT CARBOHYD 825 825
 FT CARBOHYD 903 903
 FT CARBOHYD 938 938
 SQ SEQUENCE 1070 AA; 119683 MW; CB051DF13E294603 CRC64;

Query Match 2.6%; Score 315; DB 1; Length 1070;
 Best Local Similarity 19.6%; Pred. No. 7.5e-07;
 Matches 196; Conservative 124; Mismatches 344; Indels 334; Gaps 40;

OY 5 GSGSGTQDQDAKHVDFEGQKVDHVEHGRK--NYVELKGSLSLASILGETAFYVKSQ 62
 DB 112 GGSNYMEKDG-----GDKTEERKDEHKTDSTKDNKGANNLMJLDYETSSNGOPAG 164
 OY 63 TESKYTELT---EANSKRNPCKKDGKNDVD-----RFSVKQOAGYDNKK 104
 DB 165 TLDNVLEFYTGHEGSRKN--SSNGGNPYDIDHKTTISSAIINHAFLQNTVYMKNCYTKRR 223
 OY 105 MK---CSNGM--TCAPFRRLHLCKNRPNMNSNDSSKAKHD-----LAEVCMAAYKE 152
 DB 224 RRRDMCWTKKQVCIPIDRYOLCKMELTIVNNTDTNFRDITFRKLYLKRILYDAAYE 283
 OY 153 GE-SIKTYHPKYDSKYPSPDFPMCMILARSPADIGDIIRGRPLY-LGNKKKQNKQETPR 210
 DB 284 GULLIKLNNYRIN---KDF--CKDIRMSLDGFDIIMGTMEGYSK----- 326
 OY 211 EKLQRLKEIFPKIHNDLKDEAKQRYNGDEDPNFYKLRDMWTANRETVMGAMTCSKEL 270
 DB 327 -VVENNLISIFG-----TDEKAQR-----RKQMMNESKAOIWTAMMYSVK 367
 OY 271 DNSSYFRATCDTGGQPSQTFHNKCRDKDKANAGKPRAGDGVITVPTFYDVPQYLWM 330
 DB 368 RLKGNFIWICK-----LNVAVNIEPQIYRW 392
 OY 331 FENMADRCRRKKKLEMLEKQCRKDKSDERYCSRNGYDCEQITISRGKVRMGKCTD 390
 DB 393 IREWRDVTYSELPTVQKIKKCDGKINITYDKVCK-----VPP 431
 OY 391 CFFAGSYENWIDNQRFQDKQKTYTKEISDGGGRKKRAVGSTTKY-----EGEYKS 442
 DB 432 CONAKSKYDQWITTRKKKNQWDVLS--NKFLSYVNAEKVQAGIYTPDILKQELDEPNEVA 489
 OY 443 FYEKL-KINDGYGTDAFLGL-----NNEKACKDITTDGKINFEVN----- 483
 DB 490 FENEINKRDG-----AYIELCYVSEAKKNTQEVYTVNDNAKSOATNSNFIISQVDS 544
 OY 484 -----SGGGVYGGSGGTSASGTNDENKGTFRSEYQCPDCGVQKGGQWQR 534
 DB 545 KAEKVPDSTHGNVNSGDDSTTGRAVTD-----QONGNTPA 583
 OY 535 KTVKMKMRSKLYKPPINGKAVLLLSLKLVKD-----MMILKMKKEFLTQNS----- 583
 DB 584 ESDVQR---SDIAESVSAKNVDPQKSVSKRSDPTASVTGIAAGKRNGLNASRPSSESTV 640
 OY 584 -----SDGSVSVVTGASGNSGSEKKELYDEWKYKHNVEYQVYVNGVEEEDDELKAG 638
 DB 641 EANSQDDTVNSASIPVSGEN-----PLVTYPYNGLRHSK-----DNSDSQG 682

OY 639 GCLILPNPKNKEVSEAKSQNNHADIQKTFHDFEYVVAHMLKDSIHMTKRLKSCISDG 698
 DB 683 PAESNANPDSNSKGETGKQDND-----NAKATKDS-----SNSSDG 719
 OY 699 KTMKCRNCGNKKKDCFEKVVQKQETEMKPIKDHFT-----OE 736
 DB 720 TS-----SATGDTTDAVDREINKGPE-----DRDKTVSGSKDGGEDNSJNKDAATVYGERD 770
 OY 737 GIPEGYFTTELLILKQLEKDETEENTNSIDAEEAEKLQIKLEEN----- 787
 DB 771 RIRENSAGGSTN-----DRSKNDTERKNGASTPDSKQEDATALSTRESLESTSGDRITN 825
 OY 788 ---ENNLAVYNACTEOKTLMDXLNLHNLNATCKCQOPLPEEDKSRGRSADPSDFIFPR 845
 DB 826 DTTNSLENKNGEK-----DLQKDF-----KSDTPEE-----P 857
 OY 846 EEKEDDENEDDEDEYRDE-ETAKETTESGATDTTTS 882
 DB 858 NSDQTTDAEGHDRDSIKNDKARRKRMNKDTFTKNTNS 895

RESULT 3
 RBP1_PLAYB
 ID RBP1_PLAYB STANDARD; PRT; 2869 AA.
 AC 000798;

DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE RETICULOCYTE BINDING PROTEIN 1 PRECURSOR.
 GN RBP1.

OS Plasmodium vivax (strain Belém).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=31273;

RP SEQUENCE FROM N.A.
 RA MEDLINE=9231538; PubMed=1617731;
 RX Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;

RT "A reticulocyte-binding protein complex of Plasmodium vivax
 mezoites";
 RL Cell 69:1213-1226(1992).

CC - FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO
 CC HUMAN RETICULOCYTE CELLS.

CC - SUBUNIT: HOMODIMER (POTENTIAL).
 CC - SUBCELLULAR LOCATION: MEMBRANE-BOUND.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@lsb-sib.ch).

CC EMBL: M88097; AAA29743.1; -
 DR HSSP: P36956; IAM9.

DR Malaria; Receptor; Signal; Transmembrane.
 FT SIGNAL 1 17
 FT CHAIN 18 2869
 FT DOMAIN 18 2807
 FT TRANSMEM 2808 2826
 FT DOMAIN 2827 2869
 FT SITE 1030 1032
 FT SITE 2559 2601
 SQ SEQUENCE 2869 AA; 330213 MW; B9DBE442205EBCFF CR364;

Query Match 2.5%; Score 301; DB 1; Length 2869;
 Best Local Similarity 18.1%; Pred. No. 1e-05;
 Matches 455; Conservative 343; Mismatches 890; Indels 824; Gaps 108;

OY 12 DEDAKHVDFEFOKVD-----EVHGEAKNYVELKGSLSLASILGETAFYVKSQTE-- 64
 DB 675 EEOIKELIDMKAKVHYLKELLSLKSSSVYFTENNELNTASIDYNBEGFSAKKEKADND 734

QY 65 -----SKYTE-----LIEANSKRNPCCKDG-----KGNVDVRFVSKQOAGYDN 102
Db 735 INALNSVYREDINALIEBEVEKFTENKESTLEMLKDEMEBEKLODAKETAFLKNFVSD 794
QY 103 K-----KMKCSGNCMCAPPRRLHLCKNFPNMNSNDSSAKAKDLLAEVMAAKYEBESIKT 158
Db 795 KLTDVYTMASAVINAEIKKEIKIAOKOFEN-----HKMKEFSDASTKEALON 845
QY 159 HYPKDSKYPGSDPFPMCTMLARSPADIDITRGRLYLGNKKKONGKETREKLEOKLK 218
Db 846 SMOQYNOE-----GDAL-----EKHKONSEKEBEYFKN--- 874
QY 219 EIFKIHDLNLDKBAOKRYNGDEDPNFYKLRDWMJANRETVGAMTCSKELDS----- 273
Db 875 ---ESVEEDLSRETEBEQDYTKHKNNFSRRKGE-----ISAEITMAREVINKIESOL 923
QY 274 SYFATCNDTQOGPSQTHNKCRCOKDKANAGKPRAGGDVITYPTTYDYVPQYLRMFE 333
Db 924 NY-----GYIEKYFSLIGDNEVSTAKALK-----EKIVSDSLRDKIDQYETERKE 970
QY 334 MAEDFCRKKKKLENLKQCGKDK-----SDEYRYSRNGYDCE-----OTISRK 379
Db 971 -----KTSAVENYVSTOSLSKAIDSLKRLNGSINCKKYNTDIDLRSKITLREE 1022
QY 380 GKVRMGKCTDCFPACSGYENMIDNQRKF--DKOKYTKEISDGGGRKRAVGTTKYG 438
Db 1023 VQKEPKRKDKC-----GEVTTALLKSLRDKKGIKEKINDG-----RLNSIDTKRED 1071
QY 439 YEKSFYER-----LKNDGYVDAFLGLLNNKACKDITDGGKINFEVNSGGVGGGS 493
Db 1072 LLKRYSESKSIHLKSKDQKPODP-----LNRIDEWEDI-----KRDVDELANYQYI--- 1119
QY 494 GGTSGASGTENDENKGTFRSE-----YCOPCPDQGVQHK-----GGNOMERKTKVKKM 541
Db 1120 -----SEKNVTLFEKNNSVTYIE-----AMHSIINTVAGTISNKNELLKSAVEV 1163
QY 542 R-----WKLKYPINGKAVLLK--SLKYVKDMMLIKKMKCECLTONSSDGSV 589
Db 1164 EDKLNLEQONEDYKVKVKNPEKQLEAIRGSMKLE--VINKHYSE--MPOLE---STA 1216
QY 590 SVVTTGASGSGSEK-----KELVDEMKCYKHNQYKNOGVEVED 630
Db 1217 NTLKSNKKGKENEHLELANKTKGOMRDYELKRIAELEKTEGYNELKDNENKANKEP 1276
QY 631 DDELKAGGL---CILPMPKKNKEVSEAKSONNHAD--IOKTFPHDFYVVAHMLKDSIH 685
Db 1277 EPERNITIGHVLERTVEKDKAGKYVEEMNSLKTIKLIQETSDS---ONELVTTSTI- 1331
QY 686 WTRKRLKSCISDGTMTKCRNGCNKRCDFEKWVQKQETEMKPIKDHFKTOGIEGYEYFT 745
Db 1332 --THLENAKG-----YEDVIKRNEDSIQLRBRKAK----- 1360
QY 746 TLELILKLOFLKEDTEENTENSL--DAEAEELKHLQIKILENN---LAVYNAGTE 799
Db 1361 SLETLDEMKKLVQOVNMNLGSAIQMAGISKLELNEKLVIELLSTNNSILEYVKKNS 1420
QY 800 QKTLMKDLNLHENDATCKDCPLPEEDKSRGASADPSDFIFPRPEKEDDENDDED 859
Db 1421 ESVAFSQLANGEFTKAE-----GEEKNASARLAELAE---KLKEDIVADLDYSDID 1468
QY 860 EVRDEETAKETTESATDTTSLDVCPIYGVKVLTKDNESLQDACS-----LK 907
Db 1469 KVKKIEGIRKREILMKESALT-----FWESEKFKQWSSIMENAKGKKKIE 1516
QY 908 YGGNNSRLGMKCVTPS-----GEPTTSS-----DKNGALCVPPRRRLRYIKIIV 951
Db 1517 YLKNNGGCGKANTDSOMEVGVNVSKRAHAFTVEAQVDYTKAFC-----ESIV 1566
QY 952 DMATKTES-----POASGSFASSTSGSTPPROSKLALFAVESAAIETFFL 998
Db 1567 AYVTKMNLFPNLSLMEKVKYCKEKKNDKAEKYSAKLKPYDGR---IKRAYSENERKISL 1623

QY 999 WHRRYKEERKAAVAQ--EGAGHGLPRVEEGSPRYDEPKLKE--GKIPDGLRQMFYTLGDYR 1055
Db 1624 KEKAVEKKESSQLNDVSTKSLQDNCROOLD--SVLSNIGRVKQNALQ-----YF 1673
QY 1056 DILSGSNDTTSVSKDTPRSSNDNLKNIVLLASGSTEBERKMKMYKEI--KNFRKSGTE 1113
Db 1674 D-----SADKSMKSVLPISLGAELSLDKVAAKESYKNELETQVONE 1715
QY 1114 RSAPNLVSHPOTWMENNGKYIMHGVCALTSKDKIAKVEKPKOKEPNEMIMDEANKP 1173
Db 1716 MSRLNVEGSLT-----DIDKKITDIEN-----DLKMK 1744
QY 1174 KPOYQYTNV--KLDENSGTSPTTQTOASSDNT---PTTLHPFKRPTYFRMFEWGESF 1229
Db 1745 K--GYEBGLLOKIKENADKRSNFEVLVGEINALLDPST--SIFI----- 1785
QY 1230 CREKKRLQIKVDCKVENGVGKSGDGEACDS---ISTH---DYSTYPSNCPGCG 1281
Db 1786 ---KLKLEKYMGTGLKNYGVKMNELHGEPTKSYNLETHLSNATDYSVY-----F 1833
QY 1282 KHCSYRKWIERKKIEPHKQSNAYGQOKTDATRNNGNTFDKEFCFCTLETWPDAAKFLERL 1341
Db 1834 EKAOSLRELAKREEHLRR-----REBEALFLLNDIKKYES-----LKLL 1873
QY 1342 KNGPCKTKKEYGDDIDFEKDKSTFOHTYECGPCPKFKTCNGCNGVSGLNGCDGDKS 1401
Db 1874 KEMMKVSAEYEG----- 1886
QY 1402 IDAKEIAKMSSTTDVYKRVSDNDNTNFEGBDLKADACOHANIFGIRKRDVKKCGYCV 1461
Db 1887 -----MKRDHTSVSOLVODMKTYDELTKLNDISCSVLNNAVYSYVK----- 1930
QY 1462 ICEQTNINERTDGEKYEIOLRALFRWVENFLEDYENKINDKISHCIKKGESKJINGCEKN 1521
Db 1931 VKESKHADYRDDANSMTYSMTLANV---FLSDAKIS-----SGHEFN 1971
QY 1522 SKCLEKW-----IEKKIAEMEN--IKKRPENDYENKQDPD--YVNSITL 1561
Db 1972 AEMKSNFTDLELFIYSVNSNELKKIEBDSNDVIOKEREBSQLAKDARDIYVNIKLK 2031
QY 1562 EELLIPKIVVNDQNVK-----LCVFENSKGCTLSNTQNNKENDAID 1605
Db 2032 NEFNEKLEEAANKKEVSEKVAREALKRLSOYEGIRCHENEN--RLDNTE--ELEN--- 2084
QY 1606 CMLKKLGYKAKNCGKPSGEKQSDCKEPPPLDEEDQPEENTLEPPKFCPTQPPEEK 1665
Db 2085 --LKKMTIITYD---KKSEBSGLQEM---ENEMTYNSI-----TQL 2120
QY 1666 GGETCGNKEERKD--EKKEESEPAKKEESGPAAEBPAPTAESEETETNFPPEPGTAPAP 1724
Db 2121 EGIYVSAGESKEDIKELERSNEEMRNTS---EKISTIDSKYIEMN----- 2162
QY 1725 STRAPPTDTPPLRPOADEPFDSTIIQTTIPFGVALALGSIALLFKKTKKASVGNLFQ 1784
Db 2163 ---STIDELYKLGKNCQAHWISLISY-----AMKTSK 2194
QY 1785 ILQIPLKS-----DYDIPLKSSNRYIPIVYSORLYKQTYIYMEGSDDEDKYAFMSDTT 1836
Db 2195 LIMINKKEKENEKVDYIKDNSSSTDGYETLKGFIYSGKLF-----SSASEIYQNMPTY 2249
QY 1837 DVTSESEYEELD--INDI-----YVPGSPK-----YK 1862
Db 2250 SVNPAKHEKESLNAIRDKKELLYFHQNSDISIYEGGVQNMALAYDKLNEEKREMDELYR 2309
QY 1863 TLIEVLEPESGNNTTA-----SGKNTPEDTENDIONGCIPESSKITTDNMMNOLKKEFI 1914
Db 2310 NISETKLQMEHSTDVFRPMELHKGNN---ETNNKSILLEKTKLVNDHMHSMHEAEMI 2366
QY 1915 SNMLQONPDVNDYTGNSSTNTNITTSRHNVDNNTNTMSRDNMBE-----NLL 1966
Db 2367 KNGLKATPEBQON--INNIIYSIEAEVKTLEBIDBDYGDQNOYIYEBHKQPSILLDRNAL 2445
QY 1967 LPSIHGNLYSGEEBYSYNNVKNVS-----MNDIPINBDN-----NVYSIGID 2007


```
Db 2426 MODIE-----IFKKNYNNINMEVNTETIHRVNDYIEKITNKVLQVAKTEYRQLENTKOND 2481
Oy 2008 LI-NDLSLGGKPIDIDYEVLEAKRE--NELFGTEMTKRTS-----TONVATTNSDPH 2057
Db 2482 MLONTFLKAVSIIIEYFENVKKKESILNDLYEQERLLKIGEHLEIKRNVETLSYEID 2541
Oy 2058 NOLELFHKWLDHRMCEMKKKEDILN-----KIKEENKENINNSGTYSDNKPSPH 2112
Db 2542 OKMEMSK-----MLER--KSKMMNTSYLELERANE--INNDAKQIRDDT---- 2586
Oy 2113 HVLNVDVSIQIDMDNPKTKNETNMDTNODKSTMDTILDDLKYNDPYD-----FYED 2167
Db 2587 -ILNSVLEAI-----OKRDMDAIFSOQMDADRPNPKYKAETKMMA 2628
Oy 2168 DIIVHDVDEKSSMDDIYVDHNNVTSSNMNDVPTKMHIEKNINVKKEIFE 2219
Db 2629 NEIIRQLVEKLRIGQLVQDSSEILSE-----MNSKKSALIEKE 2666

RESULT 4
ID Y67_YEAST STANDARD: PRT: 1658 AA.
AC 003661; Q04988;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE HYPOHETICAL 187.1 KDA PROTEIN IN GDA1-ERG8 INTERGENIC REGION.
GN YMR219W OR YMR261.13 OR YMR959.01.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE OF 1-711 FROM N.A.
RC STRAIN-S288C / AB972;
RA Dedman K., Brown D., Bowman S., Bartell B.G., Rajandream M.A.,
RA Walsh S.V.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 608-1648 FROM N.A.
RC STRAIN-S288C / AB972;
RA Skelton J., Churcher C.M., Bartell B.G., Rajandream M.A.,
RA Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: 249809; CAAB9934.1;
DR EMBL: 249939; CAA90190.1;
DR SGP: S0004832; YMR219W.
KM Hypothetical protein.
SQ SEQUENCE 1658 AA; 187137 MW; 3893F968305A757D CRC64;

Query Match 2 3% Score 277.5; DB 1; Length 1658;
Best Local Similarity 18.0%; Pred. No. 5,6e-05;
Matches 344; Conservative 290; Mismatches 670; Indels 605; Gaps 89;

Oy 601 SEKKELYDEMKCYKHNEVQKVNVOGEV-----EDDELKGAAGLCLLPNK----- 647
Db 2 SKKKEFFP-----RANKLKLTPRRKLKLTSLDLADESKKMDQGYSRVINDKRYRAK 57
Oy 648 -----KKEVSEAKSONNHADIQKTFHDFYVVAHMLKDS---IHWFTKRLKSCISDG 698
Db 58 PTOHSTLHESISSRSRSHI---NKSLSH-----DSARALSW---VDSLINGR 99
Oy 699 KIMKCNCGNCKKCDCEKKVVKQKTEPMK-----PINDH----- 731
```

```
Db 100 KSI--LTTLEKEDALFERKLEERQRFQJLHDSLMNKYTNQNSSHORLII LRKSQYGTDS 157
Oy 732 FKTOGIP-EGYFTTLELILKLOLEKEDTEENTENSILDAEAEELKH QKILK---LEN 787
Db 158 FQNDIEPLDSFISPL-----PDAEDESSNIDSQDEDELEH QSLIKDPDEN 207
Oy 788 E-----NMLAVNAG 797
Db 208 DVEYELSEEEKNSDGOSSPSIMILSDEYAEAGADVDVNDYAEBOY ERKNIGOEQAN 267
Oy 798 TEOKTLMDKLNLHENDAKKDCPL-----PEEDKSRG-----ISAD-----PS 837
Db 268 VENAQOISSSDSSEGOANSSEGVEMELIEDIDYESAERAKESOGABETE;SVPSKTYMQR 327
Oy 838 PD-IFPRPEEKED-----NED-----DOEDVNRDEIFAKETEGSAT 877
Db 328 TUNTKIPVIEKYESEDHKHYORSEGDARFGSVNISVDESE--DEISQAEYSANAE 384
Oy 878 D-----TTISLDYCPYVGKYLTKDNESLDACSLKGGNN-----SRGMICVTPSGEPTTS 929
Db 385 NYVHNHEHLDKELIEDIESDSSES-OSAQSESGSEDEDFYKMKNE;JSTEETENSE 443
Oy 930 SDKNGAICVPRRRRLYIKIYDMATKTESPOASGEASTS-----GS;TPPDSKALIK 985
Db 444 SKDQGF-----AKDAYTNKKEQOEENDEPEKDDIIRSLDKNFRGINKSEYSENVLE 497
Oy 986 AVEESALETFFLMHREKKEKRAVAOEGAGHGLPRVEEGSPYEDPD----- 1032
Db 498 NETDPAIVE-----RENGINDYEGYDVKGVSESLDLHESPDNL;DLAARAMLQRO 548
Oy 1033 -----KKEEKIPDGFRLQMFYTLGDRILFGSSN-----D;TYSKDTPSSS 1076
Db 549 OSRNSNCPQKEOVSESYLGHSS-----NGSNLSGRSLDESEIQILPKQFTGEN 596
Oy 1077 NDNLKNIYLLASGTFEOEREKKNKYKEIKNFRKCTERAGPMLVSHQ;VWENNGKIYH 1136
Db 597 NNNLKTDRDLSSVLEIEVEKYSKKLQDSTEKELVPLSTDTTINNS;GANEISTY 653
Oy 1137 GWYCALTSKDIKAGVEKRP-----QKLENPENLWD---EANKI;PKRPQOYTIV 1183
Db 654 ----SLDDADALISENTDVPMEIKTPKYEYVVISPSYSSTYEDNT;AMPPOVEYTS 709
Oy 1184 KIDENSGTSPRTTQOASDNTPTTILTHFVKRPTYRWMEEM-----GESEF 1229
Db 710 FM--NDPFLSLNDYEKKRDLKSTLALAA--PAFTKDAEVEAGVT;SCLTSTSGHN 765
Oy 1230 CREKKRLKQIK-VQCKVNGDVGRCSGDGEACDSISTHDYGVPSFM;PGGKHCSYR 1288
Db 766 IHTTSKETRQVSDLDDESTN-----VFENNENTDEKKNOSKN;PGV---ANSTD 812
Oy 1289 KWIERKIEFHQSNAYGOOKTDATRNNGNTEDKECKTLTETWPAK;LERLKN----- 1343
Db 813 KSTEDNTDEKYSALNY-----TNVTGDSSECDILETASVNEERL;CEROMNEAEMS 865
Oy 1344 --GPC-KTKKEYGGDDIDPEKXS-KTFQ---HTEYCGCPKFKTKQ;NGNCVGSJLGN 1395
Db 866 SGPDECYKQNDQSKTQISFTSDSPDNFQESNDNTERSS--TRYKVR-----NSD 912
Oy 1396 CDGKSIDAKELAKMSSTTDVVMYRSNDT--NTFEGDDL-----KDA;QHANIIFGIRK 1449
Db 913 LEBDESL-KKELTK-----AEVYDKLDEESSEDSTODYADPEPGNDE;SNENIYKGIK 966
Oy 1450 DWKCGYCVGVNICEQTININERTDGEYIQIRALFRWVNFLEDYNK INDRISHCIRKG 1509
Db 967 DE-----LGIVEREN-----EKVYK;VHEE--ETLFEA 991
Oy 1510 ESKSCINCEKSKSLKLEIKIAEMENIKRFNDQYENKQPOPYNVN;SILBELIPKA 1569
Db 992 NVSSSVN-----YONKDMHTDVINOBAQANYEAGRKYYIQ;TWDEE--AHIS 1037
Oy 1570 VYNDQDNVTKLVFENSKCTLISNTQNNKENDADICMLK;IGVAKN;PGRPSGKQSD 1629
```



```

Db 396 NGVTV-----PLSYNDINNALNELNSFGDLINPDYTKEPSKNIT 436
QY 1331 WPDAAKFLERLKNGPCKTNKKEYGGDIDPEKDSKTFQHTETGCPCKRTKCNCGNS 1390
Db 437 DNERKFFINEIE-KIKIEKK-----KLESCKKSE-----466
QY 1391 GLNGNCDGKSID--AKEIAKRRSSTJPDVWVRSDN-DTNFEGDGLK-----DACOHA 1441
Db 467 -----DSKSLINDITTEKEYELLINEIYD--SKFNNMIDLTFNEKMMGKRYKVEKLTJH 518
QY 1442 NIFKGIKRDVWCKGYVCGVDICEOTNINERTDGKEYIQTALFKRWVENLEFDYKNKIND 1501
Db 519 NTFASYEN-----SKHNLEKLTALKALYMEDYSLRNLYVEKELKLYKNLJSK 564
QY 1502 ISHCIRKGBGSKCINGCEKSNKCLKWKLEKTAEMENIKRRNDQYENKDDPDYVWKSIL 1561
Db 565 I-----EMEIEETLVENIKKDE--BOLFEKKITKDN-----KPEKELLEVSDIV 606
QY 1562 EELIPKRIAVNDQDNIKLCVFNESKGCJLISNTQNNKENDALIDCMKKKGVKAKNCPCK 1621
Db 607 KVOVKVILMKKIDE-----LAKTQILINKVLEK 635
QY 1622 -----PSEKOSDCKEP--PLPDEEDQNPENTLEPPKFCPTTQPEEKGETCGNK 1673
Db 636 HNHHVPSNYSKOENKQEPYLLIYLKKEID-----KLKVFMPKV-----ESLINE 678
QY 1674 EKK--KDEKKESEEPKAE-----ESCPAAEPPAPVLESEETENFPEPCT 1718
Db 679 EKKNKITBEOQSNSEPSFEGEITGOATKRPQOAGSALLEGSDVOAOOAOOAOOPVPP 738
QY 1719 GRAPPSTPAPPTPTPPLRPQADEPDPSTILOTTIPRGVALAGSIAFLKKTAKTS 1778
Db 739 VPEAKAQPPTPAP-----VNKTE--758
QY 1779 VGNLPQIOLIPKSDYDIPLKSSNRYIPVSDRYKGYIYMEGSDDEKXAFMSDTPDY 1838
Db 759 -----NVSKLDY-----LEKLYEFLNYSYCHKYILVSHSTMNKK--ILQOKYT 801
QY 1839 TSSESEYELDINDIYVPSPKYKTLIEVLEPSSNNTTASGNTPSDTRNDIONDGIP- 1897
Db 802 KEEESKLSSCD-----PLDLFNIQNN-IPV 826
QY 1898 ---SSKITDNEMNOL-----KKEFISNMLQONPDVNDYTSNGSNSTNNTTSHND 1949
Db 827 MYSMFDLSLNNLSQLPMEIYERKEMVCMYKLLKNDKIKKLEAKKVSSTVKTSSSSMQ 886
QY 1950 NNTNTTMSRDNMENMLLPISHDGNLYSGEYSYNNVNVNSMNDIPINRDNNVYSGLDI 2009
Db 887 PLSLTPROKPEVSAN-----DQTSHTNINLSLK-----LF 917
QY 2010 NDLSGKRPIDYDEVLKRKENELFGTENTKRTST-----ONVAKTNSDPIHOLELF 2063
Db 918 ENILSLGKKNKNIYOEILIGKSSSENFYEKILKDSDFYNESFNFYKSKADD-----968
QY 2064 HMLBHRDMCKMKNKEDILNKKEE-----NWK-----ENINNSGKT-----2102
Db 969 ---INSLNDESKRKKLEEDT-NKLKTKTQLSLFDLTKYKLERLFDKKTKYKGYKMOJK 1024
QY 2103 -----YNSDNKPSHNVL-NTDVSIOIDMDNFKTNEITNMJTDNDKSTMDTI 2149
Db 1025 KTLTLEQLESKLNGLNPN--KHVLONFSVFF-----NKKKEAEIA-----ETENTLE-- 1070
QY 2150 LDDLEKYNDPYIYFEDDIYHDVD---VERKSMDDIYVDHNNVTSNMADVPYTKMH--I 2204
Db 1071 -----NTKILKHYKGLVYKNGESSPLKTLSESIQTFEDNVASLENFKYLSKLEKTL 1123
QY 2205 EAMNIVNKKKEI 2215
Db 1124 KDNLEKMKKL 1134

```

```

ID MSPI_PLAFW STANDARD: PRT: 1639 AA.
AC P04933:
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE MERZOITE SURFACE PROTEIN 1 PRECURSOR (MERZOITE SURFACE ANTIGENS)
DE (PMSA) (P195).
GN MSP-1.
OS Plasmodium falciparum (isolate Wellcome).
OC Neukarya; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5848;
RN [1]
RP SEQUENCE FROM N. A.
RX MEDLINE=66014355; PubMed=2995820;
RA Holder A.A., Lockyer M.J., Odink K.G., Sandhu J.S., R-veros-Moreno V.,
RA Nicholas S.C., Hillman Y., Davey L.S., Tizard M.L.V., Schwarz R.T.,
RA Freeman R.R.;
RT "Primary structure of the precursor to the three major surface
RT antigens of Plasmodium falciparum merozoites."
RL Nature 317:270-273(1985).
RN [2]
RP REVISIONS.
RA Holder A.A.;
RL Submitted (MAR-1991) to the EMBL/Genbank/DBJ databases.
CC -I- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC (POTENTIAL).
CC -I- P1M: MERZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
CC KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC MERZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib.ch).
CC -----
DR EMBL: X02919; CA26676.1; -.
DR PIR: A24594; A24594.
DR InterPro: IPR000561; -.
DR Pfam: PF00006; EGF: 1.
DR Malaria: Merzoite; Polyprotein; Repeat; signal; Glycoprotein;
KW Transmembrane; GPI anchor.
FT SIGNAL 1..19
FT CHAIN 20..1639
FT CARBOHYD 116..116 MEROZOITE SURFACE PROTEIN 1.
FT CARBOHYD 268..268 N-LINKED (GLCNAc...) (POTENTIAL).
FT CARBOHYD 764..764 N-LINKED (GLCNAc...) (POTENTIAL).
FT CARBOHYD 768..768 N-LINKED (GLCNAc...) (POTENTIAL).
FT CARBOHYD 783..783 N-LINKED (GLCNAc...) (POTENTIAL).
FT CARBOHYD 844..844 N-LINKED (GLCNAc...) (POTENTIAL).
FT CARBOHYD 920..920 N-LINKED (GLCNAc...) (POTENTIAL).
FT CARBOHYD 964..964 N-LINKED (GLCNAc...) (POTENTIAL).
FT CARBOHYD 1058..1058 N-LINKED (GLCNAc...) (POTENTIAL).
FT CARBOHYD 1165..1165 N-LINKED (GLCNAc...) (POTENTIAL).
FT CARBOHYD 1174..1174 N-LINKED (GLCNAc...) (POTENTIAL).
FT CARBOHYD 1445..1445 N-LINKED (GLCNAc...) (POTENTIAL).
FT CARBOHYD 1526..1526 N-LINKED (GLCNAc...) (POTENTIAL).
SO SEQUENCE 1639 AA; 187618 MW; 2C255B6616C876E CR=64;

```

Query Match 2.2%; Score 265.5; DB 1; Length 1639;
 Best Local Similarity 18.1%; Pred. No. 0.00019;
 Matches 260; Conservative 189; Mismatches 457; Indels 533; Gaps 58;

```

QY 954 ARTSPQASGEASSTSGTTPPDSEKALLKAFVSAIEFFFLMHRKKEKKAVAGCG 1013
Db 61 AVTTSTPGSKGSVAGSGSVASGGSVAGSG-----SVASGC 104
QY 1014 AGHGLPRVEEGSPEDYDPEDK-----LKEGKIP-----1040
Db 105 SGNS-RTNTPSNSSDSDAKSTADLKHRVRYNLLITIKELKYQFLDITNMHMLTLDNINH 163

```

[illegible]

Qy	2022	YDEVLKRRENELEFGCEENKRTST-----QNAKTTNSPIHNOLELPHKYLDRHDDKE	2075
Db	939	YQELIGQSSSEFFYKILKDSOTFYNESSTNFKVSKADD-----INSLNDESK	986
Qy	2076	KWKNEKEDILNKLKEE-----WNK-----ENINNSGKT-----	2102
Db	987	RKLTLEDI-NLKLTKLTQLSFDLYNKKYKLTLELFPKKKTKGVKKMQIKLTKLEQLSEK	1045
Qy	2103	YNSDNKPPSHNHL-NTDVSIQIDMNPRTKNITMTMDINQDKSTMDTIIDDEKNDPRY	2161
Db	1046	LNSLNNP--KHLQNFVSFF-----NKKKEAETA-----ETENTLE-----NTKIL	1084
Qy	2162	YDFEYDDIIVHDV--VEKSSMDDIYVDHNNTVSNNMDVPFKMH-TEMNIVNNKKEI	2215
Db	1085	LKHVGLVKKYYNGESSPKLTLSSEISIQIEDNVAJSLNEFKVLSLKGKLKDNLNLEKKKL	1143
RESULT	7		
MSPL_PLAFM	MSPL_PLAFM	STANDARD:	PRT: 1701 AA.
AC	P08569:		
DT	01-AUG-1988	(Rel. 08, Created)	
DT	30-MAY-2000	(Rel. 38, Last sequence update)	
DT	30-MAY-2000	(Rel. 38, Last annotation update)	
DE	MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)		
DE	(PMMSA) (P190).		
GN	MSP-1.		
OS	Plasmodium falciparum (isolate mad20 / Papua New Guinea).		
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.		
OX	NCBI_TaxID=70153;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=88011243; PubMed=3079521;		
RA	Tanabe K., Mackay M., Goman M., Scaife J.G.;		
RT	"Allelic dimorphism in a surface antigen gene of the malaria parasite		
RL	Plasmodium falciparum";		
RL	J. MOL. BIOL. 195:273-287(1987).		
RN	[2]		
RP	REVISIONS TO 1403; 1569 AND 1629.		
RA	Tanabe K.;		
RL	submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.		
RN	[3]		
RP	SEQUENCE OF 1-115 FROM N.A.		
RX	MEDLINE=86136024; PubMed=3004972;		
RA	Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certa U.,		
RT	Stuenkelberg H., Bujard H.;		
RL	"Polymorphism of the precursor for the major surface antigens of		
RL	Plasmodium falciparum merozoites: studies at the genetic level.";		
EMBO	J. 4:3823-3829(1985).		
CC	-1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR		
CC	(POTENTIAL).;		
CC	-1- PWM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42		
CC	KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF		
CC	MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.		
CC	-----		
CC	THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION		
CC	BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION -		
CC	THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS		
CC	USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY		
CC	MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL		
CC	ENTITIES REQUIRES A LICENSE AGREEMENT (See http://www.isb-sib.ch/announce/		
CC	OR SEND AN EMAIL TO license@isb-sib.ch).		
CC	-----		
DR	EMBL; X05624; CAA29112.1; .		
DR	PIR; A26868; A26868.		
DR	PIR; B25120; B25120.		
DR	InterPro; IPR000561; .		
DR	Pfam; PF00008; EGF. 1.		
KW	Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;		
KW	Transmembrane; GPI-anchor.		
FT	SIGNAL	1	19
FT			POTENTIAL.
FT	CHAIN	20	1701
FT	CARBOHYD	110	110
FT			N-LINKED (GLCNAC. . .) (POTENTIAL).


```

RESULT      8
MSPL_PLAF  STANDARD:      PRT: 1726 AA.
ID MSPL_PLAF
AC P50495;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE (PMSA) (GPI95).
GN MSP-1.
OS Plasmodium falciparum (isolate Palo Alto / Uganda).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=57270;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99005525; PubMed=3049134;
RA Chang S.P., Kramer K.J., Yamaga K.M., Kato A., Case S.E.,
RA Siddiqui W.A.;
RT "Plasmodium falciparum: gene structure and hydropathy profile of the
RT major merozoite surface antigen (gpi95) of the Uganda-Palo Alto
RT isolate."
RL Exp. Parasitol. 67:1-11(1988).
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC (POTENTIAL).
CC -1- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
CC KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M37213; AAA29611.1; -
DR InterPro: IPR000561; -
KM Malaria; Merozoite; Polypeptin; Repeat; Signal; Glycoprotein;
KW Transmembrane; GPI-anchor.
FT CHAIN 1 19 POTENTIAL.
FT CARBOHYD 133 1726 MEROZOITE SURFACE PROTEIN 1.
FT CARBOHYD 272 272 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 567 567 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 638 638 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 827 827 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 924 924 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 944 944 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 990 990 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1016 1016 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1114 1114 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1221 1221 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1613 1613 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1726 AA; 196174 MW; 5B59CEFA2F9A026 CRC64;

```

```

Query Match 2.28; Score 264.5; DB 1; Length 1726;
Best Local Similarity 19.08; Pred. No. 0.00022;
Matches 370; Conservative 243; Mismatches 704; Indels 633; Gaps 87;

```

```

OY 444 YEAKNDGYGVDAFL--GLINNEKACKDITDGKINFKFVNSGGVGGSGSGTSGAS 500
D 25 YQELVKKLEALDAVLGTGGLFHKEMT-----LNEEETTKGASAGSGTSGTSGTS 76
OY 501 GTDENKGTFRSEYQCPDPCGQVGHKGNGWMEKRTYKKKRWKSKLVKPIGKAVLLLS 560
D 77 GTS-----GTSGTSGTSAQSGTSGTS----- 97

```

```

OY 561 LKVVKDMILKNNKKECLPTQNSSDGSVGYVTYGASG--GNSKEKLEYDEMKCYKHNEV 618
D 98 -----ASGTSGTSAQSGTSGTSGTSPSS-----RSNTL 129
OY 619 QKAVNGEVEEDDELKAGGLCLPDPKKNKEVSEAKSONNHADIOKTFHDEFFYY---- 674
D 130 PRSNTS-----SGA-----SPPADASDSDAKS---YADLKHRYVANYLFTIKEL 169
OY 675 -----WVAHMLK--DSIH-----WPKRKLKSGISGKT 700
D 170 KYPELFDLTNMLTLCIDNTHGFKYLLIDGYEINELLYKLNFEPLDLKALNDVAND--- 226
OY 701 MCRNCGNKKKCCDFEKWKOKET--EMKPIKHFKTEGIPGEYFTTLELILKLOEKE 758
D 227 -YCIIPFLKIRAMELDVLLKLVGYRRPL-DNKKDVGKMEYIKKNTTIANINELLE 284
OY 759 DTEENTENSIDAEPAELKHLQKILKLENENNLAVVNGTEQKTLMDKLNLHEINDATKC 818
D 285 GSKRTIDQNKADNDEGKKL-----YQAOYDLSTYMKOLEAHNLISYLEKRIIDTLK- 338
OY 819 KDCPLPEEDKRGSRADSPDIFIPREPEKEDENE-DDEDEVREDDETAKEETTESAT 877
D 339 -----NENIKELDLKINEIKNPPRANGPTNTLDDKKKLEEHDEKLE----- 383
OY 878 DTTSLDVCPIYGVKVLTKDNESL-ODACSLKY--GNNRSRLGMRGVTPSGEPTTSSDK-- 932
D 384 -----IATIKFNISDFTDPLELEYLREKKNKYD--VYPKSDPTKSYQIP 429
OY 933 -----NGAICVPPRRRLYIKYIDWATKTESPOASGSEASSTGSTPPPSKALLKAF 987
D 430 KVPYPNGIVLP-----LTDIHNSLAANDKKS-VGDLMDNPTEKIKIEKI 475
OY 988 VESAIEFFFLMHYK-----EKKKA-VAOEGAGHGLPRVERGSPEDYR-EDKLEGRIPD 1041
D 476 ITDKKERKIFPNNTKQIDLEFKINHTKQNKLLDEYKSKDYELLEKFEYEMKFN 535
OY 1442 GFLRQMEYTLGDYDILFSGSNDTYSK---DTPSSNDNLKNT-----VL 1085
D 536 NFDK-----DVVDKIFS-ARYTYNEKQRYNNKFPSSNSYVYVOKLKALSLYLDYS 587
OY 1086 LASSTQDERKMKKYEIKFKRCSTERSAPNLVSHPTWENNGYIWHGMYCALTSK 1145
D 588 LRKGISERD--FHHYTLKGLADIKLITEIKSEENKLTLEKNFGTLHSANASLEVY 644
OY 1146 DKIKGVK-----KPKIE-----NENLDEANKKPKPOYOYTVN-- 1183
D 645 DIVLQVOKVLLIKKIEDLKIEFLKNAQLKDSIHVPIYKPNK---PEPYLLVYLK 700
OY 1184 ---KLDE-----NSGTSPTTQTOASSDNTPTTLTHFYKRPYFRWFE 1223
D 701 KEVDKLKEFIVKMDLKKQAVLSITQPLVAASETTEDGSHSTHT-----LS 749
OY 1224 EWGSEFCREKKRL-----KQIVDCKVENGVDGRSGDGEA----- 1260
D 750 QSGTEVTEETEETEEVGHATTVTITLPRKEVY---VENSIEBKNSDQALTQTVYL 806
OY 1261 ---CDSISHTDSTYSPNFCGCGKHCSSYRKIRKIE----- 1297
D 807 KKLDELTKSYI-----CHKYIIVSNMSMOKLLEYVNLTPREENELKSCDPLDL 857
OY 1298 FHKQSN-----AYGOQTDATRNNGNTFDKE--FCKTLETWDAKFLERLK--NGPC 1346
D 858 FNIGNNIPAMYSLYDSMNNDLQHLFFELYQEMLYLHKLEENHKKLEBQOQYIGTS 917
OY 1347 KTNKEYGGDIDFEKDKSTQHTQTEYCGPCRFKYNCGNGCGVSGGLNGCNDGKSIDAKE 1406
D 918 STSSP-GNTTVN-----TAQSATH-----SNSONQSNAS----- 946
OY 1407 IAKRRSSTDVVVRVNSDNTFTFGDDLKDACQANJFKGIRKDVWKCQYVGVDCICEOT 1466
D 947 -----STNTONGAVAVSSGPAVERSHDPLVYLSISNLDG-----LYSL 986
OY 1467 NINERTDGEYIQTIAL-FKRWENPLEDYNK-INDKISHCIRKKGSGKING--CEKN 1521

```

```

Db 987 NLGNKTPVNPFLTISTTEMEKEFYENILKNNDFYNDIDIKQFVK--SNSKVITGLNETQKN 1044
OY 1522 S-----KLEKRIEKKIEMENIKRRFNDQYENK--DQPDYVKSIL---BELIPKIAV 1570
Db 1045 ALNDEIKKIKKIDTQISFLDYNNYKIKLRLFNKKKEKGODKKQIKKTLKQISKLS 1104
OY 1571 VINDQONVIR-LGVFENSKGCTLISNTQNNKENDAIIDCMKKIGVAKNKGPGPSEKQSD 1629
Db 1105 LNNPNNVLONSVFPNKKKEAEIAETENLENTK1-LLKHYKGLVKKYNGESS----- 1156
OY 1630 CKPEPPLDEEDQNEENTLEPPKFCPTTQPEEKGETCGNKE-----EKKDEK 1680
Db 1157 ---PLKTISEVSIOTEDNVANLEKFRVLISKIDKLNNDHLKKKLSFLSSGLHQLITEL 1213
OY 1681 KESEEPKAESEGPAAEPAPAESEETETNPEPPGTCGPAAPSTPADPPPD-TPEPLR 1739
Db 1214 KEVINKNYTNGSPENKKNKVEALKSYENFLPE-----AKVTYVTPPQPDVTPSPLS 1267
OY 1740 PQADPEPSTLQTTIPPGVALATGSIATFLKKTTKASVGNLPOILOIPKSDYDIPTK 1799
Db 1268 VAVSSSGSGTKEETOIP-----TSGSL-----LTLEQOVVOLQNTDEEDDSL- 1309
OY 1800 SSNRYIPIYSDRYKGYIYMEDGS-DEDKY-----AFMSDTTD--VTSESEYE-- 1846
Db 1310 -----VVLPIGSEBNDDEYLDQVYVGAISVTMDNLISGFENEYDVLY 1353
OY 1847 -----ELDINDIYVPSPKYKTLIEV-----V 1868
Db 1354 LKPLAGVYKSLKQIEKNIFETNMLNLDLNSRLKKRYFLDLESIDLQGFKHISNEYI 1413
OY 1869 LEPGNNNTTASGKNT-----PSDTRNDIQ--NDGIP-SKIKTDENQNLK--KEFIS 1915
Db 1414 IEDSKRLNLEQKNTLKSYYKIKESVENDIKFAQEGISYYEYKVLAKKYKDLIESIKVIR 1473
OY 1916 NMLQOPNDVPNDYTGNSSTNTITTSRHNVDNNTNTMSRDNMEENMLLPSIHGML 1975
Db 1474 EEKEKFPSSPP---TTPSPAKTD-----EKKKEKFLPEL----- 1506
OY 1976 VSGEESYVNNVNSMNDIPINRDNVYSGIDLINDISLGGKRPIDYDEVLRKRENELEFG 2035
Db 1507 -TNIELLYN-NLVNKKIDVLYN-----LAKKINDC-- 1534
OY 2036 TENDRTSTQ-NVATYTNNDP-IHNOLELPHKMLDRHDMCEKMKN---KEDLNLKIKEM 2091
Db 1535 --NVEDEAHVAKITKLSIDKALIDKIDLEKRNND--FDAIKRLINDDTKKMDLKGILLSTG 1590
OY 2092 NKENINNS-----GKTYNSDNKPSHNHY 2114
Db 1591 LVQNEPNTIISKLIBGKFODMNLISOHQCV 1620

RESULT 9
MSP1_PLAFC STANDARD: PRT: 1726 AA.
AC P04934:
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE MERZOITE SURFACE PROTEIN 1 PRECURSOR (MERZOITE SURFACE ANTIGENS)
DE (PMSA) (P195).
CN MSP-1.
OS Plasmodium falciparum (Isolate Camp / Malaysia).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_Taxid=5835;
RN [1]
RP SEQUENCE OF 1-1103 FROM N.A.
RX MEDLINE=86205236; PubMed=3517809;
RA Weber J.L., Leininger W.M., Lyon J.A.;
RT "Variation in the gene encoding a major merozoite surface antigen of
RL the human malaria parasite Plasmodium falciparum.";
RL Nucleic Acids Res. 14:3311-3323(1986).

```

```

RP SEQUENCE OF 1104-1726 FROM N.A.
RX MEDLINE=86143999; PubMed=3278296;
RA Weber J.L., Sim B.K.L., Lyon J.A., Wolff R.;
RT "Merozoite surface protein sequence from the Camp strain of the human
RL malaria parasite Plasmodium falciparum.";
RL Nucleic Acids Res. 16:1206-1206(1988).
CC -I- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC (POTENTIAL).
CC -I- PTM: MERZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
CC KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC MERZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage b/ and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb.ch).
CC
DR EMBL, X03831; CAN27446.1; -.
DR PIR, A23386; SAZOGM.
DR InterPro, IPR000561; -.
DR Pfam, PF00008; EGF, 1.
KW Malaria; Merozoite; Polypeptide; Repeat; Signal; Glycoprotein;
KW Transmembrane; GPI-anchor.
FT SIGNAL 1
FT CHAIN 1
FT CAROBYD 20 1726 POTENTIAL.
FT CAROBYD 133 133 MERZOITE SURFACE PROTEIN 1.
FT CAROBYD 272 272 N-LINKED (GLCNAc . . .) (POTENTIAL).
FT CAROBYD 501 501 N-LINKED (GLCNAc . . .) (POTENTIAL).
FT CAROBYD 567 567 N-LINKED (GLCNAc . . .) (POTENTIAL).
FT CAROBYD 638 638 N-LINKED (GLCNAc . . .) (POTENTIAL).
FT CAROBYD 827 827 N-LINKED (GLCNAc . . .) (POTENTIAL).
FT CAROBYD 924 924 N-LINKED (GLCNAc . . .) (POTENTIAL).
FT CAROBYD 944 944 N-LINKED (GLCNAc . . .) (POTENTIAL).
FT CAROBYD 990 990 N-LINKED (GLCNAc . . .) (POTENTIAL).
FT CAROBYD 1016 1016 N-LINKED (GLCNAc . . .) (POTENTIAL).
FT CAROBYD 1114 1114 N-LINKED (GLCNAc . . .) (POTENTIAL).
FT CAROBYD 1221 1221 N-LINKED (GLCNAc . . .) (POTENTIAL).
FT CAROBYD 1613 1613 N-LINKED (GLCNAc . . .) (POTENTIAL).
SO SEQUENCE 1726 AA; 196197 MW; DD8AD45FA352BCF3 CRK 64;

Query Match 2.2%; Score 262.5; DB 1; Length 1726;
Best Local Similarity 19.0%; Pred. No. 0.00027;
Matches 372; Conservative 243; Mismatches 695; Indels 647; Gaps 90;

OY 444 YEKLNKNDGYGVDAFL---GLNNEKACKDITDGKINKEVNSGGVYGGSGGSGSGS 500
Db 25 YQELVKKLEALDEAVLTGYGLFHKEMT-----LNDEITTYGAA:ASGTSSTGSGTS 76
OY 501 GTNDENKGFYSEYEQPCPCGVQHGKGNQWERKTKVKKMWSKLYI:PINKMVLKLS 560
Db 77 GTS-----GTSSTGTSASGTSSTGTS----- 97
OY 561 LKVVNDMLLKNMKEFCUJONSSDGSVYTTGASG--GNSSEKKE:YDEKCKYKHNEV 618
Db 98 -----ASGTSSTGTSASGTSSTGTSSTGSPSS-----RSNTL 129
OY 619 QKVVNQGVEEDDELKAGGGLCILPNPKKNKEVSAAKQNNHADIQTFPHDFY---- 674
Db 130 PPSNYS-----SGA-----SPPADASDSDAKS--YADLK IYVRYLFLTIKEL 169
OY 675 -----WVAHMLK--DS-H-----WRFKLSKISDGKT 700
Db 170 KYPELFDLTNHMLTCLDNIHGFKYLLIDGEEINELLYKLNFFDLRLK:KLVNDVACAND---- 226
OY 701 MCRMGNCNKKDCFEKRVKQKET--EMKPIKDHFKTQEGIPGYYFTTLELILKQFLKE 758
Db 227 -YQIPEFNKIRANELDVLKLVFGYRKL-DNIKDNVGMEDYIAKK NKTITIANINELIE 284
OY 759 DTBENTENSLDAEAEELKHLQIKLENNENMLAVYVNACTEDKTLMDKLNLHLENDATYGC 818

```

```

Db 285 GSKRTIDQNKADNEEGKKL-----YQAYDLSTIYNKQLEBAHNLISVLEKRIDTLTK- 338
Qy 819 KDCPLREDDKRGSRADSPDIFIPREKEDDENE--DDEDEYVADDEBTAKETTESAT 877
Db 339 -----NENIKELLDKINEIKNPPRANSNGNTPTLLDKKKKIEHEKEIKE----- 383
Qy 878 DTTTSLDVCPIYGVLTGDNESL--ODACSLKY--GGNRSRGMRVCVPSGPTTSSD- 932
Db 384 -----IAKTIKNIDSLFTDPLELEYLREKKKYD---VTPKSODPTKSVQIP 429
Qy 933 -----NGAICVPPRRRLRYIKKIYDMATKTESPOASGEASSTGSTTPPDSKALLKAF 987
Db 430 KVPYNGIYVEPLP-----LTDINSLADNDKNS--YGDLLNPPDTEKINEKI 475
Qy 988 VESAIAIEFFFLMHRK-----BEKKA--VAQEGAGHGLPRVEGSPRYD--EDLKEGKIPD 1041
Db 476 ITDNKERKIFINNLIKQIDLEKKINHTKEONKILLEDYERKSKDYELLERFYEEMKRN 535
Qy 1042 GFLQMEYTLDDYRDLIFSGSNDTTSYSK---DTPSSSNLKNL-----VL 1085
Db 536 NFDK-----DYVDKIFS--ARYIYNVEKORYNNKFSSSNNSVYNVQKLKALSTLEDYS 587
Qy 1086 LASGTEDEERKMKYKFKNPKCSTERSAPNLVSHPOTWENNGKYIWHGWCALTSK 1145
Db 588 LRKGISERD---FNHYTYLKTGLEADIKLITEIKSESENKILEKFKGLTSHANASLEVY 644
Qy 1146 DKIAKGVK-----KPKOIE-----NPNLMDENANKKKRPOYQYTNV-- 1183
Db 645 DIVLQVQKULLIKIEDLRKIEFLKNAQLKDSIHVPIYKPOK---PEPYLLYLK 700
Qy 1184 ---KLDE-----NSGTSPTTQTOASSDPTPTLLHFVRPTFYRWE 1223
Db 701 KEVDKLFKFIKVMMLKKEQAVLSSTQPLVASSETTEDGSHHT-----LS 749
Qy 1224 EWGSPFCERKKRL-----KOIYDCKVENGVDYRCSGDGEA----- 1260
Db 750 QSGETEVEETEETEETVHTTWTTLPLREKVV---VENSIEHKSDNSQALTKTYVL 806
Qy 1261 ---CDSISHTDSTVPSFPCGCGHSCSYRKWIEKKIE----- 1297
Db 807 KKLDEPLTKSYI-----CHYIILVSNSSMOOKLEEVNLPPEENELKSCDPLDL 857
Qy 1298 FHKOSN-----AYGOQKTDATRNNGNTFDE---FCCTLETPWDAKFLRLK--NGPC 1346
Db 858 FNIGNNIPAWYSIVSDSMNNDLQHLFFELYOKEMITYLHKLEBNHKKILLEEQOIIGTS 917
Qy 1347 KTNKEYGGDIDFEKDSKTFQHTYCGCPKPKFNCGNGCGVSGLNGCDGKSIDAKE 1406
Db 918 STSSP-GMTYVN-----TAQSATH-----SNSQNOQSNAS----- 946
Qy 1407 IAKRRSSTVDVVMRYSDNDTTFEGDDLKDACQAHANIFKGRKDWKCGYCVGDICROT 1466
Db 947 -----STYTQNGVAVSSGPVAVESHDBPLVLSISNDLKG-----IYSL 986
Qy 1467 NINERTDQKEYIQIRAL-FKRWFENFLEDYNN--INDKISHCIRKKGSGKJNG--CEKN 1521
Db 987 NLGKKTVPNPLJLSTTEMEKFEYENILKNNDTYFNDDIKQYOVK--SNKRYITGLTERQKN 1044
Qy 1522 S-----KLEWIEKKAEMENIKKRFNDQYENK---DQPDYNVKSII---EELIPKIAV 1570
Db 1045 ALNDEIKKLKJTLQLSFDLYKKYKILDLRLFNKKKELODMQKQIKLILKEQLESKLNS 1104
Qy 1571 VNDDQNIYK--LCVFENSGCGLLISNTQNNKENDAIMDKLKLGYAKAKCPKPSGKOSD 1629
Db 1105 LNNHNHVLQNFVSFPNKKKEAIELETETLENTYI--LTKHYKGLVKKYNGESS----- 1156
Qy 1630 CKEPPLPDEDDQNPENTLTPPKFCPTTPTOPPEKGGETGKNKE--EKK----- 1677
Db 1157 ---PLKTLSEVSIQTEEDYANLEKF-----RVLKSIDKLDNHLGKKKLSFLSSGLIH 1208
Qy 1678 --DEKKESEEPKAEESGPAAEPAETAESEETETNPPEPPGTPAAPPSTPAPPTD-T 1734

```

```

Db 1209 LITELKEVYKNNKNTYGNSENKKNVNEALKSYENFLPE-----AKVTYVTPPPQDVT 1262
Qy 1735 PPLRPQADPEPDSITLIQTTIPFGVALAGSIALFLKKTAKASVGNLFQLOLPKDYD 1794
Db 1263 PPSLVSFVRSVSGSGSTKETQIP-----TSGSL-----LTLEQOVVOLQNDYEE 1305
Qy 1795 IPTLKSNNRIPYVSDRYKRYIYMEGDS--DEDKY-----AFMSDTPD--VTSSESE 1844
Db 1306 DDSL-----VLPFGESEDNDDEYLDQVYTGAIYSTMONIISGFENE 1348
Qy 1845 YE-----ELDINDIYVPSPKYKTLIEV----- 1867
Db 1349 YDVIYILKPLACVYRSLLKQIEKNIFETNLINDINSRLKRYKFFLYLDESLLQFHHIS 1408
Qy 1868 ---VLEPSGNNTTASGKNT-----PSDTRADIO--NGCIP--SKITDNEMNOLK-- 1910
Db 1409 SNEYIIEDSFKLNSBQKNITLAKSYIKYESVENDIKFAQGISIYERKVLAKYODLESI 1468
Qy 1911 KEFISNMLQNDPNDVNDYTSNGNSTNTNTTTSRHHVNDNTNTWSRDNMEENILLPSI 1970
Db 1469 KKVIKKEKEKFPSSPP---TTPSPAKTD-----EOKKESKFLPFL 1506
Qy 1971 HDGNLYSGEEYTVVNNVNSKNDIPINDNNVYSGIDLINDLSGCRPIDIYDEYLKKE 2030
Db 1507 ---TNIEETLVN--NLVKNKIDYILN-----LKAKI 1531
Qy 2031 NELFGTEKRTSTQ--NVAKTNSDPIHNOLELPHKMLDRHRDM--CEKKMN--KEDIL 2084
Db 1532 NDC-----NVEKDEAHVKITKLSDLKALIDDKIDL----KHNDEAIKLLINDTKRDM 1583
Qy 2085 NKLKEENKKNINNS-----GKTYSNDKPSHHV 2114
Db 1584 GKLLSTGLVQNFPMNTIISKLEGFQDMLNLSHQCV 1620

```

RESULT 10

HRX_HUMAN

AC 003164: 014845: 016364: 013743: 013744: 020MA3;

DT 01-OCT-1993 (Rel. 27, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 01-OCT-2000 (Rel. 40, Last annotation update)

DE ZINC FINGER PROTEIN HRX (ALF-1) (TRITHORAX-LIKE PROTEIN).

GN MLL OR HRX OR ALL1 OR TRX1 OR HTRX.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

OX NCBI_TaxId=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=93046667; PubMed=1423624;

RA Tkachuk D.C., Kohler S., Cleary M.L.;

RT "Involvement of a homolog of Drosophila trithorax by 11q23

RT chromosomal translocations in acute leukemias.";

RL Cell 71:691-700(1992).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=96290553; PubMed=8703835;

RA Nilson I., Loechner K., Slegler G., Greil J., Beck J.D., Fey G.H.,

RA Marschalek R.;

RT "Exon/intron structure of the human ALF-1 (MLL) gene involved in

RT translocations to chromosomal region 11q23 and acute leukemias.";

RL Br. J. Haematol. 93:966-972(1996).

RN [3]

RP SEQUENCE OF 1-1909 FROM N.A.

RX MEDLINE=93390935; PubMed=8378076;

RA Yamamoto K., Seto M., Komatsu H., Iida S., Akao Y., Kojima S.,

RA Kodera Y., Nakazawa S., Ariyoshi Y., Takahashi T., Ueda R.;

RT "Two distinct portions of UFG19/ENL at 19p13 are involved in t(11;19)

RT leukemia.";

RL Oncogene 8:2617-2625(1993).

RN [4]

RP SEQUENCE OF 1317-2328 FROM N.A.

RC TISSUE=Brain; PubMed-1303259;
 RX MEDLINE-93265134; PubMed-1303259;
 RA Djibail M., Sellert L., Parry P., Bower M., Young B.D., Evans G.A.;
 RT "A trichorax-like gene is interrupted by chromosome 11q23
 translocations in acute leukaemias.";
 RL Nat. Genet. 2:113-118(1992).
 [5]
 RN SEQUENCE OF 1251-1538 FROM N.A.
 RX MEDLINE-94215165; PubMed-8162575;
 RA Gu Y., Alder H., Nakamura T., Schlichtman S.A., Prasad R., Canaan O.,
 Saito H., Croce C.M., Canaan E.;
 RT "Sequence analysis of the breakpoint cluster region in the ALL-1 gene
 involved in acute leukemia.";
 RL Cancer Res. 54:2326-2330(1994).
 [6]
 RN SEQUENCE OF 1251-1654 FROM N.A. (ISOFORM 14P-18B).
 RX MEDLINE-95315013; PubMed-7598802;
 RA Mbangkollo D., Burnett R., McCabe N., Thirman M., Gill H., Yu H.,
 Rowley J.D., Diaz M.O.;
 RT "The human MLL gene: nucleotide sequence, homology to the Drosophila
 trz zinc-finger domain, and alternative splicing.";
 RL DNA Cell Biol. 14:475-483(1995).
 [7]
 RN SEQUENCE OF 1212-1603 FROM N.A.
 RX MEDLINE-95315013; PubMed-7794749;
 RA Marschalek R., Grell J., Lochner K., Nilson I., Slegler G.,
 Zweckbronner I., Beck J.D., Fey G.H.;
 RT "Molecular analysis of the chromosomal breakpoint and fusion
 transcripts in the acute lymphoblastic SEM cell line with chromosomal
 translocation t(4;11).";
 RL Br. J. Haematol. 90:308-320(1995).
 [8]
 RN SEQUENCE OF 1421-1540 FROM N.A.
 RX MEDLINE-94020842; PubMed-8414518;
 RA Forster A., Rabblits T.H.;
 RT "A method for identifying genes within yeast artificial chromosomes:
 application to isolation of MLL fusion cDNAs from acute leukaemia
 translocations.";
 RL Oncogene 8:3157-3160(1993).
 [9]
 RN CHROMOSOMAL TRANSLOCATION WITH GAS7.
 RP MEDLINE-20183971; PubMed-10706619;
 RA Megonigal M.D., Cheung N.-K.V., Rappaport E.F., Nowell P.C.,
 Wilson R.B., Jones D.H., Addya K., Leonard D.G.B., Kushner B.H.,
 Williams T.M., Lange B.J., Felix C.A.;
 RT "Detection of leukemia-associated MLL-GAS7 translocation early during
 chemotherapy with DNA topoisomerase II inhibitors.";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:2814-2819(2000).
 CC -I- FUNCTION: POSSIBLY ACTS AS A TRANSCRIPTIONAL REGULATORY FACTOR.
 CC -I- SUBCELLULAR LOCATION: NUCLEAR.
 CC -I- TISSUE SPECIFICITY: HEART, LUNG, BRAIN AND T AND B LYMPHOCYTES.
 CC -I- DISEASE: INVOLVED IN ACUTE LEUKEMIAS BY CHROMOSOMAL TRANSLOCATIONS
 T(11;19)(Q23;P13.3) THAT INVOLVES MLL AND MLLT1/ENL;
 T(11;19)(Q23;P13.3) THAT INVOLVES MLL AND MLLT2/AF4; T(9;11)(P22;Q23)
 T(4;11)(Q21;Q23) THAT INVOLVES MLL AND MLLT2/AF4; T(9;11)(P22;Q23)
 THAT INVOLVES MLL AND MLLT3/AF9; T(6;11)(Q27;Q23) THAT INVOLVES
 MLL AND MLLT4/AF6; T(11;17)(Q23;Q21) THAT INVOLVES MLL AND
 MLLT6/AF17; T(X;11)(Q13;Q23) THAT INVOLVES MLL AND MLLT7/AFX1;
 T(10;11)(P12;Q23) THAT INVOLVES MLL AND MLLT10/AF10;
 T(1;11)(Q21;Q23) THAT INVOLVES MLL AND AF10; T(11;19)(Q23;P13.3)
 THAT INVOLVES MLL AND ENL; AND T(11;19)(Q23;P23) THAT INVOLVES MLL
 AND GAS7.
 CC -I- SIMILARITY: BELONGS TO THE TRITHORAX FAMILY OF TRANSCRIPTION
 FACTORS.
 CC -I- SIMILARITY: CONTAINS 1 BROMODOMAIN.
 CC -I- SIMILARITY: CONTAINS 1 SET DOMAIN.
 CC -I- SIMILARITY: CONTAINS 3 PHD ZINC-FINGER DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as the content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)

CC or send an email to license@isb-sib.ch.
 CC -----
 DR EMBL; L04284; AAA58669.1; -;
 DR EMBL; 269744; CAA93625.1; -;
 DR EMBL; 269745; CAA93625.1; JOINED.
 DR EMBL; 269746; CAA93625.1; JOINED.
 DR EMBL; 269747; CAA93625.1; JOINED.
 DR EMBL; 269748; CAA93625.1; JOINED.
 DR EMBL; 269749; CAA93625.1; JOINED.
 DR EMBL; 269750; CAA93625.1; JOINED.
 DR EMBL; 269751; CAA93625.1; JOINED.
 DR EMBL; 269752; CAA93625.1; JOINED.
 DR EMBL; 269753; CAA93625.1; JOINED.
 DR EMBL; 269754; CAA93625.1; JOINED.
 DR EMBL; 269755; CAA93625.1; JOINED.
 DR EMBL; 269756; CAA93625.1; JOINED.
 DR EMBL; 269757; CAA93625.1; JOINED.
 DR EMBL; 269758; CAA93625.1; JOINED.
 DR EMBL; 269759; CAA93625.1; JOINED.
 DR EMBL; 269760; CAA93625.1; JOINED.
 DR EMBL; 269761; CAA93625.1; JOINED.
 DR EMBL; 269762; CAA93625.1; JOINED.
 DR EMBL; 269763; CAA93625.1; JOINED.
 DR EMBL; 269764; CAA93625.1; JOINED.
 DR EMBL; 269765; CAA93625.1; JOINED.
 DR EMBL; 269766; CAA93625.1; JOINED.
 DR EMBL; 269767; CAA93625.1; JOINED.
 DR EMBL; 269768; CAA93625.1; JOINED.
 DR EMBL; 269769; CAA93625.1; JOINED.
 DR EMBL; 269770; CAA93625.1; JOINED.
 DR EMBL; 269772; CAA93625.1; JOINED.
 DR EMBL; 269773; CAA93625.1; JOINED.
 DR EMBL; 269774; CAA93625.1; JOINED.
 DR EMBL; 269775; CAA93625.1; JOINED.
 DR EMBL; 269776; CAA93625.1; JOINED.
 DR EMBL; 269777; CAA93625.1; JOINED.
 DR EMBL; 269778; CAA93625.1; JOINED.
 DR EMBL; 269779; CAA93625.1; JOINED.
 DR EMBL; 269780; CAA93625.1; JOINED.
 DR EMBL; D14540; BAA03407.1; -;
 DR EMBL; L01986; AAA92511.1; -;
 DR EMBL; U04737; AAA18644.1; -;
 DR EMBL; S78570; AAB34770.1; -;
 DR EMBL; X83604; CAA58584.1; -;
 DR EMBL; S66432; AAB28545.1; -;
 DR EMBL; AF231998; AAC26332.2; ALT_TERM.
 DR TRANSFAC; T02337; -;
 DR MIM; 159555; -;
 DR InterPro; IPR001214; -;
 DR InterPro; IPR001487; -;
 DR InterPro; IPR001965; -;
 DR InterPro; IPR002857; -;
 DR Pfam; PF00628; PHD; 3.
 DR Pfam; PF00856; SET; 1.
 DR Pfam; PF02008; ZF-CXXC; 1.
 DR PROSITE; PS50014; BROMODOMAIN_2; 1.
 DR PROSITE; PS50280; SET; 1.
 KW Proto-oncogene; Chromosomal translocation; DNA-binding; Bromodomain;
 KW Nuclear protein; Zinc-finger; Metal-binding; Transcription regulation;
 KW Alternative splicing.
 FT DOMAIN 17 102
 FT DNA_BIND 169 180 ALA/GLY/SER-RICH.
 FT DNA_BIND 217 227 A.T HOOK (BY SIMILARITY).
 FT DNA_BIND 301 309 A.T HOOK (BY SIMILARITY).
 FT ZN_FING 1431 1482 A.T HOOK (BY SIMILARITY).
 FT ZN_FING 1484 1533 PHD-TYPE 1.
 FT ZN_FING 1566 1627 PHD-TYPE 2.
 FT DOMAIN 1703 1748 PHD-TYPE 3.
 FT DOMAIN 3840 3969 BROMODOMAIN (DIVERGENT).
 FT SET.
 FT DOMAIN 137 143 POLY-GLY.
 FT DOMAIN 561 564 POLY-PRO.
 FT DOMAIN 568 571 POLY-PRO.
 FT SITE 1444 1445 BREAKPOINT FOR TRANSLOCATION TO FORM MLL-

FT	VARSPILC	1407	1444	GAST ONCOGENE.
FT	CONFLICT	144	144	MISSING (IN ISOFORM 14P-18B).
FT	CONFLICT	317	379	E -> ELTQTQPCSWRTGHHDKKTEPEFRLAMSWCUN (IN REF. 2).
FT	CONFLICT			GLIIISELEKPPQVNRKDEKGPPLKKEKTYVROSPRPIK VRIIPSSKTRDTAKTIKOLLORA -> VESLIINMKSPKSG KTRRHHLKQRTIOLSDKALEGLISQLGLFLLQKQOMPLL SNTSIQO (IN REF. 1).
FT	CONFLICT	556	556	Q -> E (IN REF. 2).
FT	CONFLICT	1487	1487	R -> G (IN REF. 5).
FT	CONFLICT	1603	1603	S -> SCGE (IN REF. 2 AND 7).
FT	CONFLICT	1616	1616	S -> C (IN REF. 6).
FT	CONFLICT	1937	1937	Q -> H (IN REF. 4).
FT	CONFLICT	2181	2181	P -> S (IN REF. 4).
FT	CONFLICT	3718	3718	R -> G (IN REF. 2).
FT	CONFLICT	3759	3759	N -> D (IN REF. 2).
FT	CONFLICT	3813	3813	D -> G (IN REF. 2).
FT	CONFLICT	3901	3901	R -> A (IN REF. 2).
Query Match 2.2%; Score 262.5; DB 1: Length 3969;				
Best Local Similarity 18.2%; Pred. NO. 0.0074;				
Matches 450; Conservative 292; Mismatches 888; Indels 837; Gaps 112;				
QY	10	MODEBAKHYLDEFGGVHD-----VHOEAKNYVELEK-----GSLSLASTIG	52	
DB	859	SKDRDQKVEYEDKSHREPEREKRRSRREKRRKGEISQSSALYPVGNVSEKVVG	918	
QY	53	ETAFYVKSQTESKYTELIEANSKRNPCKDKGNDV-----DRFSYKE-----OAGYDN	102	
DB	919	EDVAVSSSAK-----KATGRKSSSHD--SGDITDSVTLGDTTAVATKLLIKGRGN	968	
QY	103	-KKMCSNMTAPF-RRLHLNKNFPNNNSDSSKAKHDLAENCMAKYGESIKTHY	160	
DB	969	LEKTMLDLPTAPSLDEKTEKTL-----LSTPSSSVYKHSSTSSISMLQAD-----	1014	
QY	161	PYDSSVYPSDSDPMSCTMLARSPADIGDIIRGDIYLGNNKKKONKEPEREKL	217	
DB	1015	-----KLPTMDRVASLTKAKAQLQLEKSSLSKOTDQPKAO--QGESSESSEFSVRGPI	1068	
QY	218	KEIFKIHNDLKDEAKRYNGDEPNFYKFLREDWMTARFVWGMATCSKELDNSYPR	277	
DB	1069	KHYCRAAAVALORKRA---VPPDDMPTLSAL--PW--BEEREKILSSMGND--DKSSING	1118	
QY	278	A-----TCNDTGCGSPQTHNK-----CRCDKDKGANAG---KPK	308	
DB	1119	SEDAEPLAPIRKIPVTRNNKAPOEPRPVYKGRSRRCGCPGCOVPEDCGYCTNCLDKPK	1178	
QY	309	AGDDGVTVIPTYEDVPOYLWPFEEAEADFCFKKKKKLENLKOCGRGKKSDBEYHYCSN	368	
DB	1179	FGG-----RNIKKQCCMKRKNCONLDMPKSP	1203	
QY	369	GVDCEQTTISRKGRVNRMGKCTDCFPACGSYENWIDNORKQFQKOKYUTEISDG-----	422	
DB	1204	AVLQOAKAVKKKKKKK-----TSEKKDKDESSVYKNNVDDSQKPPR	1246	
QY	423	GGRKKRANGCTTKYBESYEKSFYEKLKNDGYGVDAFLGLLINNEKACKDITDGGKINFEV	482	
DB	1247	SAREDPAPKSSSEPPRPKRYVEESBE---GNVSA-----PEGESQATTPASRKSRSQY	1298	
QY	483	NAGGGVGGGSGSTSGASTNDENKGTFRSELYCOPRCPOGHHNGGNGMEKRTYVKKKR	542	
DB	1299	SQPALVYI-PRPPRTGPR-RKEVPRKTTSEPRKKORP---PESGPEQSKKKVAPRS	1352	
QY	543	WSKLYPIKNGKVVLLLSLKVVKDMIMLLKNNKEFCLTQNSSDGSVGVSVTTGASGANS	602	
DB	1353	IPVKRKP-----KEKKRPPVKNQENAGLNLITST--LSNGNS	1389	
QY	603	KKELYDEMCKYHNEQYKVNVOGEVEEDDELKAGAGLCLIPN-ERKNKEVSEAKSONNH	661	
DB	1390	KOKI-----RADGVNIRVYDFEKDECEAENYEMWGAGLILSVPLTRPVVCFCLASSG	1442	
QY	662	ADIQ-----KTFHDFEYTVVANHMLKDSI-HNRTIRLKSCLSDGK-----TMCSR	704	

Dd	1443	VEFYCYOVCSEPFHFKFLENERPDLBOLENNCCBRCKCYHCGROHATKOLLECNKR	1502
Qy	705	NGCNKC-----DCEKYNKOK-----	726
Dd	1503	NSYHECJGPNYPKPKTKKKKWMJCTKCVKCKSCGSTTPGKMDAMSHDFSLCHDAKL	1562
Qy	727	-----PIKD-----HFTQSGIPGYYFTTLELKLQPLK	757
Dd	1563	FAKGNFCPLCKCYDDDDYESKMMQCKDQWVHSCENLSDMEYI-----LSMLP	1614
Qy	758	EDTEENENSLDAEAELELKHOKIKLENNUL--VYVNAGETQKTLMDKLNHELNAT	816
Dd	1615	ESVAYTCYNKTERHPAEMWRALEKEIOLSLKOYLTRPLHNSRTTSHL-----RYQAA	1667
Qy	817	KCKDCPLREEDKSGRSADSPDIFIPREPEKEDDENEDDEDEVDDETAETKETTESGA	876
Dd	1668	KPRLNPEETESIRSRSSPGPPRPVLTVESKQDO-----QPLDEGVYKMKDQCN	1720
Qy	877	TDTTT--SLDVCPIYKVLTYTDNDSLQDASLKYGGNNSRLGRCYTPSGEPTTSDKGA	935
Dd	1721	TSVLEFSDDIYKIIQAAINSD-----GGQ-----PEIKANSWVKSFF	1758
Qy	936	ICVPRRRRLYIKIIVYMAKTESPQASGEASTGSTPTPDSKEMALLKAFVESAEIET	995
Dd	1759	I-----ROMEVFPWFVSYKKSRTMEPKVSSNSG-----MLPNVLPSPSLDN	1801
Qy	996	FELMHRYKEE-----KRAY-AOEGAGHLP-----RVEGSPXY	1028
Dd	1802	YAOQEREENSHTBQRPMLKIIIPAKPKGSGPDPPLHPPTPLSTDRSREDSPEL	1861
Qy	1029	DPEDKLKGGKIPDGFLOMYTL-----GDYRLIFSGSDTTSV-----SKDTPSS	1076
Dd	1862	NPPREIDN-----ROCALCUTYGDDSANAGRLLYIGONEMTHVNCALMSAEVEDD	1914
Qy	1077	NDNLKNIYLLASGSTOEERKMKK-----YKEIKNF-----RK--C	1110
Dd	1915	DGSLKNVMAVIRKQJLRCEFCOKPGATVGCCLTSTSYHFMCSBRANKCYFLDDKKYVC	1974
Qy	1111	STERSAPNLVSHPOTWENNGKIYHGMVC--ALTSKDKIAKGYEKKRQKIEPNUL-	1165
Dd	1975	QRNRDLKGEVVRP-----NGFEVFRVAFVDEGJLSLRKPLNGLE-----PENIM	2021
Qy	1166	-----WDEANKKRPPOUYOTYVNLKDENG-----TSP	1193
Dd	2022	MIGSMITDCLGILNDSDCEDKLFPIGYOCSRYWSTTARRCVYTCKIVECRRPVER	2081
Qy	1194	RTTOQASDNTPTTLNHFUKRPTRYFWMFEMGESCRKKRYLOI---KYDCVGENG	1250
Dd	2082	DINSTVEHDEN--RTIHH---SPTST-----ESSKESQNTAELISPPSDRPHSQT	2130
Qy	1251	VGRCSGDG--EACDSISTHDYSTVSPFNCPC-----	1280
Dd	2131	SGSCYVNIHISKVPIRTPYS--PTQRPCRCRLPSAGSPRTTHEIYVGDPLLSGLR	2188
Qy	1281	-----GKHCSTRKRIETKKLEF---HKQSNAYGQCTDTRNNGTFFDKFCKTILETPD	1333
Dd	2189	SIGSRHSTSLSS--PQSKSLRIMPMKMTGNTYSRNVSSVSTTGATBLE-----S	2238
Qy	1334	AAKFLERLKNRCKPTKNEYGDGDDIDPEKQSKTPQNHIEYCGPCPKFKTCQNGNGCSGLN	1393
Dd	2239	SAKVVDHVL--GPLNSSLISLG-----QNT-----STSSNLQRIYVYVGKN	2277
Qy	1394	GNCGDKSIDAKELAKMRSSTTVVNRKVSNDNTNTEGDLK-----DACOHANIFKG	1446
Dd	2278	SHLDGSSSEMKQ-----SSASDLVSK-----SSLKKGKTYLVLSKSSBSAHNVAVPG	2327
Qy	1447	IRK-----DVMKGVGVGADICEOTNINERTDGKEYIQIRALFRKRVENFLE	1493
Dd	2328	IPKLAPOVHNTSRHELNVSKIG-----SFAEPSVS--FSSKEALSPHLHLRGORDND	2380
Qy	1494	DY-----KINDK-----ISHCIKKGBSKINCENKNSK--CLEKWEK	1551
Dd	2381	QNTDSTQSANSSPDEDTEVYTKLSCGMSNRSSITENHMSSSRDRORQKSKSCETEPEK	2440


```

Db 492 ELLEKYEEMKNNNDK-----DYYDKIFS-PAITYNVEKORYNNKSSNNSSVYVQ 543
Oy 1084 -----VLLASSTEOEREMKYYKELKRNKRCSTERSAPNLVSHQTMWENNGK 1132
Db 544 KLKALSLLEYSLRKIGISED---FNHYTLKGLLEADIKKLEIKSEKILIEKKF 600
Oy 1133 YIMHGMWALTSKDKIAKVER-----KPKIE-----NPEMLDEANKK 1172
Db 601 GLTHSANSLSLEVDIVKLOVOKVILLIKKIEPLRKIEFLKNAQKDSJHVNIYRPOK- 659
Oy 1173 PKPPOYQYTNV-----KLDE----- 1187
Db 660 ---PEPYLLYLKKEVDLKEFIPKVKMDLKEQAVLSITOPLYAASETTEGGHSTHT 716
Oy 1188 ---NSGSPRTQTOASSONTPTTLHFAKRPYFPMFEMGESFCREKKRLKOIKVDC 1245
Db 717 LSGOGETEVTETEVEETETVCHTTVTITLPP-----KEESAPEKVKV--- 759
Oy 1246 VENGDVGRCSGDGA-----CDSISTHDYSTVPFNCPCGKHCSSYRKWIERKKI 1296
Db 760 VENSIEHKSNDNSQALTKTYLKLKDEFLTKSYI-----CHKYILVNSSMDOKLL 810
Oy 1297 E-----FKQSN-----AYGOQKTDATRNNGNTFDKE---FCK 1326
Db 811 EYVNLTPREENELKSCDPLDLFNIONNIPAMYSLYDSMNIDLQHLFELYOKEMITYLH 870
Oy 1327 TLETWPDAAKFLERLK---NGPCFKNKKEVGDDIDEEKSKTFPHNTEYCGPCPKFTNCON 1384
Db 871 KLKEBNHKLLEBKOKQITGTSSTSSP-GNTTVN-----TAQSAKH-----SNSON 915
Oy 1385 GNCVSGINGNCDDKSIDAKEIAKMRSSITDVYKRVSDNNTFEGDDLKDACOHANIF 1444
Db 916 QOSNAS-----STNONGVAVSSGPAYVESHDPLTVLSTISNDL 954
Oy 1445 KGIRKDWKCGYVCGVDICEPTNINERTDGEKEYIOIAL-FKRWVNFLEDYNK-INDKI 1502
Db 955 KG-----IVSLNLNGNKTKVNPPLISTTEKEKFEYENLKKNDDYFNDI 999
Oy 1503 SHCJIKGSGKING---CEKNS-----KCLEMKTEKIAEMENKRRPNQYENK---D 1551
Db 1000 KOYFK---SNSAVITGLTQKALNDELKIKLKDITQLSFDLYNKYKLLDLRFNKKKELG 1057
Oy 1552 QPDYVNSITL---ELLIPKIAVVDODNVIK-LCVFENSKCCTLTISNTONKKENDAIOM 1607
Db 1058 QDKQIKKLTLLKEOLESEKLSLNPNHVLQNFVSFFKKKKEAEALAEENTLENKI--L 1115
Oy 1608 LKIKGVAKKNCPRKPSGKQSDCKEPPPLPDEEDONPEENTLEPPKFCPTTQPEEKGG 1667
Db 1116 LKHVKGVLKYYNGESS-----PLKTLSEVSIQTEDNVANLEK-----RALSKIDG 1161
Oy 1668 ETCGNKE---EKK-----DEKKESEPAKESGPALEAPATAESEGETENFP 1713
Db 1162 KLNDLHLGKKKLSFLSSGLHLITELKEVTKNNKYNNGSENKKNKYNAKLSYENFLP 1221
Oy 1714 EPPGTGPAAPSTPAPPTPD-TPPPLRPOADEPDSITLQTTIFGVALAGSIAFLFLK 1772
Db 1222 E-----AKVTTVYTPPQDVTPSPLSVYSGSGSGSTKEEQIOP-----TGSGL----- 1264
Oy 1773 KKTASVGNLFQIIQIIPKSDYDIPTLKSSNRYIPYVSDRYGKTYIWEQDS-DEKX- 1829
Db 1265 -----LTELQOVVOLQNYDEEDDSL-----VLPJFGSESDNDEYLD 1301
Oy 1830 -----AFMSDITD---VYSSESEYE-----ELDINDIYYPG 1857
Db 1302 QVVTGEAIVSMNDILSGFENEYDYLKPLAGYVRSIAKQIEKNITTFNLDINDISR 1361
Oy 1858 SPKYKTLIEV-----VLEPSGNMTASGKNT-----PSDTRNDIO- 1892
Db 1362 LKKKKYFLDVLESDLMQFKHSSMEYIIEESFKLNSQKNTLLAKSYKIESEVENDIKF 1421
Oy 1893 -NDGIP--SSKITDNEWMOLK--KEFISNMLQNOPDVPNDYTSNGSSNTNTITTSRHNV 1948

```

```

Db 1422 AOEGISYEVKYLAKYKDDLESIKKIVKEKEKFPSSPP---TTPPSPAKTD----- 1469
Oy 1949 DNNTWTMSRDMNEENLLPSIHONLTSGEYSYNNVMVNSMDIPIN-----RDNNV- 2002
Db 1470 -----EOKRESKLPPL-----TNIETLYN-MLVKIDYILINLAKKINDCAVE 1512
Oy 2003 -----YSGIDLINDLSGGKPIDYD-----EVLKRENELEFGTENTKRTS-----T 2044
Db 1513 KDEAHVKITKLSDLKAIDOKIDLFKNTNDPFAIKKLIN-----DOTKMDLGLSLTGLV 1567
Oy 2045 QNVAKTNSDPDIIHNDLEFHKWLDHRHDMCEKMKNKED-----ILNKIKE 2089
Db 1568 QNFPNTIISKLEIEG---FQDMLNISOHQCVKQKOPENSGCFRLHDERECKCLLN-YKQ 1623
Oy 2090 EWNKENINSGKTYNSDN 2107
Db 1624 EGDK-CVENPMPCTNENN 1640

RESULT 12
ID ATRX_MOUSE STANDARD; PRT: 2476 AA.
AC 061687;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE TRANSCRIPTIONAL REGULATOR ATRX (X-LINKED NUCLEAR PROTEIN)
DE (HETEROCHROMATIN PROTEIN 2) (HP1 ALPHA-INTERACTING PROTEIN) (HP1-BP38
DE PROTEIN).
GN ATRX OR XNP OR HP1BP2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98213653; PubMed=9545503;
RA Picketts D.J., Tastian A.O., Higgs D.R., Gibbons R.J.;
RT "Comparison of the human and murine Atrx gene identifies highly
RT conserved, functionally important domains.";
RL Mamm. Genome 9:400-403(1998).
RN [2]
RP SEQUENCE OF 325-1176 FROM N.A.
RX MEDLINE=97133299; PubMed=9978696;
RA le Douarin B., Nielsen A.L., Garnier J.-M., Ichinose H.,
RA Jeanmougin F., Losson R., Chambon P.;
RT "A possible involvement of TIF1 alpha and TIF1 beta in the epigenetic
RT control of transcription by nuclear receptors.";
RL EMBO J. 15:6701-6715(1996).
CC -1- FUNCTION: COULD BE A GLOBAL TRANSCRIPTIONAL REGULATOR. MODIFIES
CC GENE EXPRESSION BY AFFECTING CHROMATIN.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
CC -1- SIMILARITY: CONTAINS 1 PHD-FINGER DOMAIN.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AF026032; AAC08741.1; -.
DR EMBL; X99643; CA67962.1; -.
DR MGD; MGI:103067; Xnp.
DR InterPro; IPR000330; -.
DR InterPro; IPR001650; -.
DR Pfam; PF00176; SNF2_N; 1.
DR Pfam; PF00271; helicase_C; 1.
KW DNA repair; Nuclear protein; DNA-binding; Helicase; ATP-binding.
FT DOMAIN 219 267 PHD-FINGER.
FT NP_BIND 1579 1586 ATP (POTENTIAL).

```

Query Match 2.1%; Score 251.5; DB 1; Length 2476;
Best Local Similarity 17.9%; Pred. No. 0.0013;
Matches 435; Conservative 325; Mismatches 831; Indels 835; Gaps 115;

51 LGPTAFYKVS-----MOTESKYTELLEANSKRPNCKKDGKNDVDFSVKEDQAGYDNKKM 105
102 VNEKATENSENDITWQSLPKGTIVQPEPVLNEDKDEKGE--FRSRKMKADNLIK 158
106 KCSNGM---TCA-----PPRLHLCNKNPNMNSDSKAK----- 138
159 RGEDGLHGVSTACGQOVNHFQKDSIYRHPSLKVLICNCFKYYMSDISDSDGMDQ 218
139 -----HDLAEVCAAAKYGESIKTHYPRKYDSKY-----PGSDPPMCTM 177
219 CRMCAEGNLIICDPCFNHAFCKKILRNLRGKELSTIMDENQWVCYICQPEPLDLVTA 278
178 IARSPADIGDITRGADLYGNKKKONGKETEREKLEQKLKELFKIHNLKDEKQAKRY 237
279 CNSVENIEQLLQ-----QNRKKIKVDEKTSKYCDQTSK-----FSPKKSSSSC 323
238 NGDEDPNFKLEEDMWTANRETVMGAMCSKELDNSSYFRATCNPTGGPSQTHNKCRC 297
324 NGEER---KLEE-----SCSGSV-SSTY-----SHSLASLP 350
296 KDKGANAGCPKAGDDVTIVPYFDYVPOLYLWMEFEMAD-----FCRRK--KKKLE 347
351 KMIKKTKK-----LIETTSNMNNSYIKFLQAOADNSEMTSAMKLCQLKSPFSYLD 401
348 NLEKO--CGRKKSDER-----YCSRNGYDCEQITSRKGVKRMKGCITDCEPACGSYEN 400
402 DIRKHALEEDLNEIEQLADVHEKENTKDLKST--DAKSETKLKG-----EK 449
401 WIDNORKOPDK--OKKYTEISDGGGRKKRAGVGTTKYEG--YEKSYEKLKNDGYGTVA 457
450 SYSTEREFKLDAHSVKAIDGEOBRAHKSTSGEHKSGRKDGSOYEPTNT---PEDL 505
458 FLGLLNNEKAC-KDITDGGKINFKEVNSGGVYGGSGGT-----SGASGTIDENKGFY 511
506 DMDIYVSPSPVEDLFDLSLE-SAMEYQSSADYQGDGNSGTPEPELESSYKLVNVSADS- 562
512 RSEYQPCPCDCVQHKGQNMQRKTKVKKMRWSKL-----YKPLNG----- 552
563 -----RGNIKSKVTAKVRKELFVKLIPVLSNPSPIKGVCOEVSQEKNGR 607
553 -----KNVLLKLSLKVYKDMMI-----LKKMKKEFCITLQNSSD 585
608 KSSGVARSEKCRPREEISDHENNVITLLESDDLRSPPVKTPTLLKROTESPANMSNDE 667
586 GSVGVSVTTGASGNSSEKKELYDEMCKY---KHNEVQKVN--VOGEVEEDDELKAGAG 639
668 ESNQTKKEKQKSGPIRKDKDRNSADCATDNPKPHKVPRAKQPVIGDQSDSEM----- 722
640 LCILPPKKNKKEVSE-AKSQNNHADIQKTFHDFEYIYVAHMLKDSITHMFKRLKSCISG 698
723 LNVL-----KEASQWGHSSSDTDINE-----PQNNHNG 751
699 KTMKCRNGCNKCKDCFEKKNVKKQTEPMKPIKDHFKTOEGIPRGYVTTLELILKIQFLKE 758

752 KTGKDDNGKRK-----RKNSTS-----GSDFDTKKG-----KSTELSLSK-----KK 789
759 DTEENTENSLDEBEAEELKHLOKILKLENNENMLAVYNAGTEBOKTLMDKILNHELDAIKC 818
790 RQNSSESSWYDELEIREITMSRI-----GAARKSV----- 820
819 KCCPLPEEDKSGRSADSPDFFIPRPEEKEDDENEDDE--DEVIRD--EETAKETEGS 875
821 -----PEKKEEDSEDEQKRVYDNGCHEAKTKTOEGS 854
876 AADYTTSLDVCPIGVKVLTKDNESIQ-DACSLKYGG--NNSRLGMRCVYTGSEPTTSSDN 933
855 SADD-----TGDEGRQGGSCSIAGSGIEKVRGVEFERELICRPVYSSD-- 898
934 GAICVPPRRRLYIKKIYDMATKTESPOASGSEASSTSGTTPPDSKELLAAPFESAII 993
899 GA-----EKPSYKEBNVNS-----PEDKR-----YSKKE 923
994 ETEFLMHRKYEKKKAVAOAGAGHGLPRVEGSPEDPEDKLEKGIPLDKFLRQMFYTLGD 1053
924 KTKHLRSROSKRGKGGSSDGT-----DREPKQESDESSEGEKKQ;-RQRETKG- 972
1054 YRDILFSGSNDTYSKDTPPSSSNDLKNIVLLASGSTEQ-----IEKMKYKEI- 1104
973 -----KRAPDLKGETLKR-----EQWDSSSDG;ERLPEEBEIEP 1007
1105 --KNFRCKSTERSAPNLVSHPQTMWENNGKIYHGNVCAITSKDKIAKIVEKPKQIENP 1162
1008 FSKGIKQSKTDPAG-----GEKKGKK--WKDKSC--EKKEELSD;YVRLPKGSDSC 1054
1163 ENLMDPANKKPPPOXYQYTNVVLDSNS-----GTSPTQTQOASSNT;YTLTLHPVKRT 1217
1055 DSEEDKTR-----NRVSLREKKRPSLPAKSP--GNRPCSSSDT;KSLK----- 1097
1218 YRFMEWGESFCREKRRKIKQIVDCKYENGVDVCGSGDGACDSISYHDSVPSFNC 1277
1098 -----QGC-CDSEKRPKRIDL-----RERRNSS;RNNKVEKYSAS 1133
1278 PCGKHSSYRRIWIEKKILEFHQSNAYGOQKTDATRNNGTFDEKEFC;YLETWPDAAKF 1337
1134 SSDAEGSSBDNK--KQK-----QRTSAKKKGTGNKREK--R;SLRATP----- 1173
1338 LERLKNPGKTKKEVGGDDIDFEKDSKTPQHTIEYCGPCPKFTNQN-----GNCQVS 1390
1174 -RRQVDITSSSDIGDDQNSAGESSDQ-----KIRPVTEMLVPSHTGCGSS 1224
1391 G-----LNGNCGDGSIDAKETAKMRSSSTDVVMVSNNDYTFGGDDLKDCQHA 1441
1225 GDEALSKSPATVDDDDNDNPEBNRIAK--KMLLEIKALSSDEGS--SDEPDG----- 1277
1442 NIFKGIKRDVWKCGYVGVNDICEQTNINERTDG--KEYIOIRALEKRAVENFLEBYNKI 1498
1278 ---GGKRRIGK-----QSESPADGDGLRQGLAVANQV-----NSESDSDE 1316
1499 NDK-----ISHCIKKEGSKCINGCEKNSKLEKWIETKIAEBENIKKFNDOYEK 1550
1317 ESKRPYRRLRLRHKLTLSDE--SGEKPTPKP--EHKEAKRNRKRVSS--EDS 1366
1551 DQPDYVWKSILEELPKIYVNDODNVIKLCVFNESKGTCLISNTQNKENNAIDCMCLK 1610
1367 EDTDPOESVSEE--VSESEDBQRPRT--RSKKAKLEBNQSYQOK-----KR 1412
1611 LGVAKANCPGKPSGKQSDCKEPRPLPRDEDONPREENTLEPRFCPTTQPREKGGETC 1670
1413 RRIKYQD--DSSSENKSHSEEDKKEGDEDEDEDE-----DENDSKSP 1458
1671 GNKEEK-----KDEK--KESEEPRAKEESGPAABEPRAPESEETETNFPERPOTGAAPR 1724
1459 GGRKKIRKILDDDKLRTTQNALKEE--ERRKRIAREREREREKILREVLEIEDASP- 1513
1725 STPAPPTPDTPLPRQAQEPFDSTLTQTFPGVAL--ALGSIATLF-----LKKKKR 1776
1514 -TKCPRTTLVLVDENETKEP-----LVQVNRMMVILKLNQVDQVQFMRDCCSESVKIK 1568

OY	1777	ASVGN-----	:	:	: :	LQIIIDIPSPDVIDPLTSS--NRPIYPYSDDKKCT	1816
Dd	1569	KSPSGGCIILAHCMGLKGTLDVVFSFLHTVLCLCKLDPSTALVCPLNTALAMNNEFK---	:	:	: :	----	16355
OY	1817	IYMEGSDDEDKYA FMSDTTVDTSSESEX---EELDINDIYVPGSPKYTLIEVLEPSG	:	:	:	18739	
Dd	1626	--WQGIENDNEKELEVSELATAVKRPPQRSYWLDMWGDSGVMIIGEVYNRL-----	:	:	:	16744	
OY	1874	NNITASGRKNPISDRNDIQDGFI----PSKITDNEMNLCKREFLSNMLOPNPDVPDY	:	:	:	19299	
Dd	1675	---AGRRNVKSRLKDIFKAVALVDGPPEFVDC-----GHLLKMVASVSXRM	:	:	:	1720	
OY	1930	TGSNSTVTNTITTTSRHNVDNNNTMTSRIHDNMEENLLPSIHGMLYSGEEYSRVNW/-	:	:	:	19888	
Dd	1721	NMISRIRILLTGPR-----LNQLNI-----EHCHVNANIK	:	:	:	1751	
OY	1989	-NSMNPIPNDNNVYSGIDLINDSLGGKPDIYDEVLRKENELF---GTENTKRTS	:	:	:	20433	
Dd	1752	ENLGSIKIKEFRNFREIN--PIQGQCADSTMVDV-RVMKRAHILIYEMLAGCVQRK---	:	:	:	18033	
OY	2044	TQNVAKTINTSDPIHN-----GLELFHKWLD-	:	:	:	20659	
Dd	1804	-DYALLKFELRPHEYLAVRMTAIOCKLYOYYDLHLTVGSTEGRGAKACALKFDFOQ	:	:	:	18623	
OY	2070	--HRDMCEKWKNMEDILNKLEEKNENINSNGTKYSNDRKSHNHVLTIVSQIDMDN	:	:	:	21277	
Dd	1863	MLSIMTHHPICOLDIYISKRNKGTFPEDSMIDERIASDS-----ETSKSLSSD-EK	:	:	:	19122	
OY	2128	PKTANEITNMDFTNODKSTMQTIJLDLEKRYNDPYFYDEFEDLIYHDVEK------	:	:	:	2178	
Dd	1913	KKPKGKKGXKXS SSSSGSGSD-----ADV EIAVKWNRSRSGC	:	:	:	1948	
OY	2179	--SSMODIYDVHHNWTSSNNNDVRPTKM	:	:	:	2202	
Dd	1949	GCGMMDD-----TGNNPVSCLK	:	:	:	1966	
<hr/>							
RESULT_13							
ID	TPR_HUMAN	STANDARD;				PRT; 2349 AA.	
AC	p12270;						
Dt	01-OCT-1989 (Rel. 12, Created)						
Dt	01-OCT-1996 (Rel. 34, Last sequence update)						
Dt	01-OCT-2000 (Rel. 40, Last annotation update)						
De	NUCLEOPROTEIN TPR.						
OS	Homo sapiens (Human).						
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.						
OX	NCBI_Taxid=9606;						
RN	[1]						
RP	SEQUENCE FROM N.A.						
RX	MEDLINE=93064711; PubMed=1437155;						
RA	Mitchell P.J., Cooper C.S.;						
RT	"The human tpr gene encodes a protein of 2094 amino acids that has extensive coiled-coil regions and an acidic C-terminal domain.";						
RL	oncogene 7:2329-2333(1992).						
RN	[2]						
RP	REVIEWS, AND CHARACTERIZATION.						
RX	MEDLINE=95096166; PubMed=7798308;						
RA	Bryd D.A., Sweet D.J., Panté N., Konstantinov K.N., Guan T., Septhre A.C.S., Mitchell P.J., Cooper C.S., Aepli U., Geran L.;						
RT	"tpr, a large coiled coil protein whose amino terminus is involved in activation of oncogenic kinases, is localized to the cytoplasmic surface of the nuclear pore complex.";						
RL	J. Cell Biol. 127:1515-1526(1994).						
RN	[3]						
RP	SEQUENCE OF 1-142 FROM N.A.						
RX	MEDLINE=88262257; Pubmed=3387099;						
RA	King H.W.S., Tempest P.R., Merrifield K.R., Rance A.J.;						
RT	"tpr homologues activate met and raf."						
RL	Oncogene 2:617-619(1988).						

Query Match	Best Local Similarity	Matches 390;	Conservative 340;	Mismatches 771;	Indels 815;	Gaps 97;
38 VSELGSLA---	2.1%.	Score 249.5;	DB 1;	Length 2349;		
92 LTEKKLELIEDNRALIAQSOTRKEELFAKRLIRTNELSG-ELEYLEDVKRL--	16.8%.	Pred. No. 0.0013;				
95 KEQAGYDNKKMKSCNGMTCAFPRRLHLCNKNFNNMNSNDSSAKAHDLAELVMAAKYEGE						
149 -----NEKLKESN-----						
155 SIKTHPYKDYSPSDPMPCTMLARSA---DIDDIIRGRDLVGNKK-----KKQN 204						
182 RLDEKKELLHSQNTWMLNLTETKTDELLALGKREKENETLELCKNLNNKEEVSRLDEQWN 241						
205 GKETEPEKLEQKLEIFPKIHDLNKDEKA--OKRPNGED-----PNFYKTLREDWMTANR 257						
242 GLKTSNEHLQKHVEDLTKLKE-AKEQASAMEEKHNELMAHIKISNLTK-----290						
258 ETVWGAMTCKELDNSSYFRATCNDTGGPSQITHNKKCRCKDKGAMNAGPKVAGDGVTV 317						
291 -----SAADS-----EAKSNELTRAVEELHKLK-----EAGEANKAI- 324						
318 PTYEDYVQYLRFWEEMAEDECRKKKKKLEMLLEKQCRGKDSDEYRYSRNGYDCEQITIS 377						
325 -----QDLLEVEQSQSDOMEKLEIKIGRLEKEL-----EVANDLLSATK 364						

OY	378	RKGR-----	-----VNRKGCCTDCFA-----	-----CGSYN-----	WID 403
Db	365	RKGAIISEBELAAMSPTAAAVAKIVPRGMATITELVAAVETODOLLEKLETKENRKINKYID			424
OY	404	NOROFQOKKQYTPTEISDGGRRKRAVVG--TTEVEEYKSPFEKLIKNOGYVDAFLGL 462			
Db	425	EIVAVEKAPILKROREPERAKOAAVASISVLEBAMEI--ORLOED-----			471
OY	463	NNEKA-----	-----CKDITDGGKINFKVNS--GGGVVG-----	SGSGTSAS 500	
Db	472	-TDKANNOSVLEBNDNRMIQVCKDLSQQRVLMLLEBARCNHVIKREDEVSALISSSS 530			
OY	501	GTNDENKGTFRSEYCO-----	PCDDCGVHKGNOMERTKYKKRWMSLYKPIWG 552		
Db	531	EVISQHLVSYNIEELOQONRLLVALRELGETFREREDQETTSKITELO--LKLESALT--			588
OY	553	KMVLILKS-----	LKVYDMILKKNKKEPITON-----		582
Db	569	ELEBOLRSROHOMOLV--DSIVRODMYRILLISOTGVAILPLHASSLDVSLASTPKRPST 647			
OY	583	-----SSDGSVYVTTGASGSEKKELYDEMCKYKHNVOKVYNOGVEEEDDELKAG 638			
Db	648	SOYSTARPVYIESTAIETAKKALKOLOEIPENTYKKEAENKIONOLE-----			698
OY	639	GLCILPMPKKNEVSEAKSONNHADIOKTFHDEFFYYVAHMLKD-----	SIHMT 688		
Db	699	-----KLQEQYTDLRSGNTKISTQIDPASKRY-----	EMLODNEGYRREITSLHEIN 746		
OY	669	KRLKSC-----	ISDGTMKCRNGCKKCCDFKWKV--OKETEMKPIKDHKTO-----	735	
Db	747	OKLATYQOKEOIINTMTODLR--GANEKLAVALAVRAENIKKEKEMIKLESEVLISOORESL 805			
OY	736	-----EGIPREGYVTTLELLKLOPLKEDEEENTSLA--EAE--ELKHLOKILKLENE--	788		
Db	806	LAEBORGONLLITNIQTT--OGILENSETETKORLSQJLEKLEHEISHLKV--KLENEVE 860			
OY	789	-----NNLAV-----	VAAGTEOKTLMDKLNLHETLNATKCK 819		
Db	861	ORHUTLRNLDOYQLDTRKQJODFTENLHLNLTKEKLNAGKEIATFLKHLSNMEVQVAS--	917		
OY	820	DCPLPEEDKSKGNADSPDIFIPREBEKDDENEDDEVEDERODEETAKETTESATDT 879			
Db	918	-----QSSORTGKG-----	OPSNKED--VDLVSOLROTEGEYVNDLKEKELKST 959		
OY	880	T-----	TSLDVCPYIGVKVLTKDQESJODACSLKYGGNNSRLGMRVTPSGPRTSS 930		
Db	960	SNVEOYQAMTSLSE-----	ESLNKEQYTE--		985
OY	931	DKNGAICVPPRRRLYIKKIVDATKTESPOASGSEASTSGSTPPDCKEALLKAFVES 990			
Db	986	-----VRKNI-----	EVRLK--ES 997		
OY	991	AAIETFLMHRYKEEKAVAOEGAGHGLPVEEGSEPYEDPKLEKGIPTDGLROMFYT 1056			
Db	998	AEFOYT-----	OLEKK--LMEVEKEKOELO--DDKRR--	IESMEQ 1032	
OY	1051	LGDRDILFSGSD-----	TTSVSKOPPS-----	SSNDLKNIVYLASG 1089	
Db	1033	LSEKLTSSVONVQALORASTASLNEQOARRDCOEQAKIATVEAQONKYTERELMLHAD 1092			
OY	1090	-----STEOERKANKKEIKNFKCSTERSAPNLVSHQTMWENNGKXITMGWCAITLSKDK 1147			
Db	1093	VEALOAKKEQSKMAVSRQHLEFTQKAEBSOLLECKASMBERRML-----	KDE 1141		
OY	1148	IAGVY-----	EKKPQKTEHP--ENLMDANKKRPPOYOTYTNVTKLDENSGTSPRTTQOJA 1200		
Db	1142	VSKVCRCRCEDEKONRLHJOIEKLSKVAAVSEKVOGPLANVSLSE-----	EG 1190		
OY	1201	SSDNTPTTLTHFVPRPYTFPMFMEWGSPFCREKKRKLKLOIKYDCKYKENDVARGSGDGA 1266			
Db	1191	KSQOEILFIRRE-----	KEIAET--REFAOVESLRTRORE-----	1229	
OY	1261	CDISITHSYSTVSPFNCPGGCKHCSSYRWKIEKKITFEFKOSNAYOQOCTDATRNGNTF 1320			

[illegible]

RX MEDLINE-9235776; Pubmed-1496004;
RA Adams J.H., Sim B.K., Dolan S.A., Fang X., Kaslow D.C., Miller L.H.;
RT "A family of erythrocyte binding proteins of malaria parasites.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:7085-7089(1992).
CC -1- FUNCTION: BINDS TO THE HUMAN ERYTHROCYTES DUFFY BLOOD GROUP
CC DETERMINANT.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- SIMILARITY: HIGH, TO P.VIVAX DUFFY RECEPTOR.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb.ch).
CC -----
DR EMBL, M90694; AAA29603.1; -
KW Malaria; Receptor; Glycoprotein; Signal; Transmembrane;
KW Multigene family.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 1153 DUFFY RECEPTOR, BETA FORM.
FT DOMAIN 22 1085 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1086 1106 POTENTIAL.
FT DOMAIN 1107 1153 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 134 134 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 179 179 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 202 202 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 348 348 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 467 467 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 626 626 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 722 722 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 847 847 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 856 856 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 900 900 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 910 910 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 935 935 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1153 AA; 130471 MW; 6497BD16CE7BEC01 CRC64;

Query Match 2.1%; Score 249; DB 1; Length 1153;
Best Local Similarity 19.0%; Pred. No. 0.00064;
Matches 197; Conservative 125; Mismatches 343; Indels 370; Gaps 45;

QY 14 DAKHVLDFGOKVHDEYHG-----AKNYVS---ELKGSLSLASILGETAFYV 58
DB 112 EAQNVAD--NNYIDDKANGKYEKTNTKNGEGARMVWLDYDISGHPDGLID---NV 165
QY 59 KSMQTESKTYLLEIANSRNCKKDKGNDVDRESYKQ--AGIDNKKM-----KSN 109
DB 166 VELGTED-----GSELENSSK--GQDHYRNNRKEKMSGVYNQETLQKNVRRNN 216
QY 110 GM-----TCAPFRRLHLCKNKPNN-----NSNDSKAKHDLAEVMA 148
DB 217 KRRKGRDMDQPTKKDKVCPDRRYOLCKELTILVNNKTHSHDITFLKLNLKELTYD 276
QY 149 AKYGESIKTHYPRYDSKYPSGDPFPMCTMLARSPADIGDITRGADLY--LGKKKKRQNGE 207
DB 277 AAVGDDLTL--LKKRYNNVYSED--LCKDKIMSLSEDFDITIMGTDMEGIGYS----- 322
QY 208 TERKLEOKLKEIKFKKHIDNKLKDEAKRYNGDDEPNRYKLRBDMWNTNRETVGAMTCS 267
DB 323 ---OVENNLTLVFTGTGYQLD-----RKKVMWESKRYIWEATLLS 361
QY 268 KELDNSSYFRATCDMGQSPQTHNKCRCDKGKAGNAGKPKAGGDVYITVTFYDVPQY 327
DB 362 VKKRLNGYSANMCKE-----DVOI-----NVEPQI 386
QY 328 LRMFEENAEDFCRRKKKKLLENLEKQCRGKDKSDERYCSRNGYDCEQTISSKRGVRRMGKG 387

DB 387 RWRIREMGDMYSELPRKQRIKKECDKRLYYTLNRICITMS-----P 428
QY 388 CTDFEFACSGYENWIDNORQFQKQKTYKEISDGGKKRAVGTTKYE-----GYE 440
DB 429 CND---SKLYDQMITTKKKQWDVLS--TKFSYKKGQIITENTITAYILDKQELNGFN 483
QY 441 KSFYE-----KLR-----NDGY 452
DB 484 EVMFENEINKRDYNYIDICLCADEPNKNTQEHKLKLSAPKLETFORSHSTIQPMSSGA 543
QY 453 GTVDAFLGLLN-NEKACKDITDGGKINFEVNSGCGVGGSGGTSCASGT-NDENKGT 510
DB 544 EKVQGDLAHGININAAVKSTTD-----EAKKBDGONGNTVAESNIKGDNIENEA 596
QY 511 YRSEYQCPRCPCGVQHGKGNQWERTKVKYKRWKSLKYPINGKWLKSL---KVVYD 566
DB 597 NVDTY-KVYERSADTKRA-----TDITTEGEKRLNTSYSGSEITVYKINIPDGDYKD 649
QY 567 MMI-----LKRNMKEFCLTONSSDQSVGSVYTTGASGNSBKELYDEMKCYKHN-- 616
DB 650.VSAAVENSENPLETKHKIFEPDSKNDNSSENS-----GSM-----EFKATSSMPI 694
QY 617 --EVQKVVQGEVEDEDELKAGAGCLTLPNPKKNKEYSEAKSONNADIDKTHDFEY 674
DB 695 TEAVESSSAEQVOVEDSAHRSVNTG-----RDNSTISAATSD----- 731
QY 675 WVAHMLKDSIMHRTKRLKSCISDQKTKMCRNGKCKKCCFEKWKQKETEKKPI----- 728
DB 732 -----DGLSSGDKRRESLTS-----IENADGSDPYQGSILMN 763
QY 729 -----KDFKTOEG-----IPGYYPTLELILKLOFLKEDTEEN 763
DB 764 LLNDPSVAGGSGKSHIKTEENEGSOAELDKGNVOIAEQRATTEVQPERPDLSDTQGN 823
QY 764 TENSIDAE-----AELKHLQKLIK-----LENENNLAVNAGTEQXTLMDKL 807
DB 824 VPRSGNKNKNEGATALLGSAESLESNESYHKTIDNTTHGLENK-----NGNEKQFOKHF 877
QY 808 LNHE-LNDATCKCKCPLEEDKSGRSADSPDFFIRPEKEDEDEDND--DDEDEVRD 864
DB 878 MNMDMLNDQTS-SPQTSDDQTSNQTSSDQTS-----NOTSSDQTSDDQISSDQTS 930
QY 865 EETAKETTESGATDT 879
DB 931 QTSNQTSSDQTIPT 945

RESULT 15
ATTRX_HUMAN
ID ATTRX_HUMAN STANDARD; PRT; 2375 AA.
AC P46100; P51068; O15886;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE TRANSCRIPTIONAL REGULATOR ATTRX (X-LINKED HELICASE II) (X-LINKED
DE NUCLEAR PROTEIN) (XNP).
GN ATTRX OR RAD54L OR XR2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_Taxid:9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-97123494; Pubmed-8968741;
RA Pickelits D.J., Higgs D.R., Bachoo S., Blake D.J., Quarrell O.W.,
RT "ATTRX encodes a novel member of the SNF2 family of proteins: mutations
RT point to a common mechanism underlying the Attr-X syndrome.";
RL Hum. Mol. Genet. 5:1899-1907(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-97386582; Pubmed-9244431;
RA Villard L., Lossi A.M., Cardoso C., Proud V., Chitroni P.,

RA Coliaux L., Schwartz C., Fontes M.;
RT "Determination of the genomic structure of the XNP/ATRX gene encoding
RT a potential zinc finger helicase."; Genomics 43:149-155(1997).
RL [3]
RN SEQUENCE OF 743-2375 FROM N.A.
RX MEDLINE-9517911; PubMed-7874112;
RA Stayton C.L., Dabovic B., Gulsano M., Gecz J., Broccoli V.,
RA Giovannazzi S., Bossolasco M., Monaco L., Rastan S., Boncinelli E.,
RA Bianchi M.E., Consalez G.G.;
RT "Cloning and characterization of a new human Xq13 gene, encoding a
RT putative helicase."; Hum. Mol. Genet. 3:1957-1964(1994).
RL [4]
RN PRELIMINARY PARTIAL SEQUENCE FROM N.A.
RX MEDLINE-94214473; PubMed-8162050;
RA Gecz J., Pollard H., Consalez G., Villard L., Stayton C.,
RA Millaesau P., Khrestchatsky M., Fontes M.;
RT "Cloning and expression of the murine homologue of a putative human
RT X-linked nuclear protein gene closely linked to PKX1 in Xq13.3."; Hum. Mol. Genet. 3:39-44(1994).
RL [5]
RN SEQUENCE OF 2284-2375 FROM N.A., AND VARIANTS ATR-X.
RX MEDLINE-95211835; PubMed-7697714;
RA Gibbons R.J., Picketts D.J., Villard L., Higgs D.R.;
RT "Mutations in a putative global transcriptional regulator cause X-
RT linked mental retardation with alpha-thalassemia (ATR-X syndrome)."; Cell 80:837-845(1995).
RL [6]
RN VARIANTS ATR-X.
RX MEDLINE-97467722; PubMed-9326911;
RA Gibbons R.J., Bachoo S., Picketts D.J., Aftimos S., Aisenbauer B.,
RA Bergoffen J., Berry S.A., Dahl N., Fryer A., Keppler K., Kurosawa K.,
RA Levin M.L., Masuno M., Nerl G., Pierpont M.E., Slaney S.F.,
RA Higgs D.R.;
RT "Mutations in transcriptional regulator ATRX establish the functional
RT significance of a PHD-like domain."; Nat. Genet. 17:146-148(1997).
RL [7]
RN VARIANT JM GLN-2014.
RX MEDLINE-96224392; PubMed-8630485;
RA Villard L., Gecz J., Mattel J.-F., Fontes M., Saugier-Verber P.,
RA Munnich A., Lyonnet S.;
RT "XNP mutation in a large family with Juberger-Marsidi syndrome."; Nat. Genet. 12:359-360(1996).
RL -1- FUNCTION: COULD BE A GLOBAL TRANSCRIPTIONAL REGULATOR. MODIFIES
CC GENE EXPRESSION BY AFFECTING CHROMATIN. MAY BE INVOLVED IN
CC BRAIN DEVELOPMENT AND FACIAL MORPHOGENESIS.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- DISEASE: DEFECTS IN ATRX ARE THE CAUSE OF X-LINKED ALPHA-
CC THALASSEMIA/MENTAL RETARDATION SYNDROME (ALSO KNOWN AS ATR-X
CC SYNDROME). ATR-X IS AN X-LINKED DISORDER COMPRISING SEVERE
CC PSYCHOMOTOR RETARDATION, CHARACTERISTIC FACIAL FEATURES, GENITAL
CC ABNORMALITIES, AND ALPHA-THALASSEMIA.
CC -1- DISEASE: DEFECTS IN ATRX ARE THE CAUSE OF JUBERGER-MARSIDI SYNDROME
CC (JM). JM IS A RARE X-LINKED RECESSIVE DISEASE CHARACTERIZED BY
CC SEVERE MENTAL RETARDATION, GROWTH FAILURE, SENSORINEURAL DEAFNESS,
CC MICROGENITALISM AND EARLY DEATH.
CC -1- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
CC -1- SIMILARITY: CONTAINS 1 PHD-FINGER DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, U72938; AAB49971.1; -;
DR EMBL, U72935; AAB40699.1; -;
DR EMBL, U72904; AAB40699.1; JOINED.
DR EMBL, U72907; AAB40699.1; JOINED.

DR EMBL, U72908; AAB40699.1; JOINED.
DR EMBL, U72909; AAB40699.1; JOINED.
DR EMBL, U72910; AAB40699.1; JOINED.
DR EMBL, U72911; AAB40699.1; JOINED.
DR EMBL, U72912; AAB40699.1; JOINED.
DR EMBL, U72913; AAB40699.1; JOINED.
DR EMBL, U72914; AAB40699.1; JOINED.
DR EMBL, U72915; AAB40699.1; JOINED.
DR EMBL, U72916; AAB40699.1; JOINED.
DR EMBL, U72918; AAB40699.1; JOINED.
DR EMBL, U72919; AAB40699.1; JOINED.
DR EMBL, U72920; AAB40699.1; JOINED.
DR EMBL, U72921; AAB40699.1; JOINED.
DR EMBL, U72922; AAB40699.1; JOINED.
DR EMBL, U72923; AAB40699.1; JOINED.
DR EMBL, U72924; AAB40699.1; JOINED.
DR EMBL, U72925; AAB40699.1; JOINED.
DR EMBL, U72926; AAB40699.1; JOINED.
DR EMBL, U72927; AAB40699.1; JOINED.
DR EMBL, U72928; AAB40699.1; JOINED.
DR EMBL, U72929; AAB40699.1; JOINED.
DR EMBL, U72930; AAB40699.1; JOINED.
DR EMBL, U72931; AAB40699.1; JOINED.
DR EMBL, U72932; AAB40699.1; JOINED.
DR EMBL, U72933; AAB40699.1; JOINED.
DR EMBL, U72934; AAB40699.1; JOINED.
DR EMBL, U72935; AAB40699.1; JOINED.
DR EMBL, U72936; AAB40699.1; JOINED.
DR EMBL, U72937; AAB40700.1; JOINED.
DR EMBL, U72938; AAB40700.1; JOINED.
DR EMBL, U72939; AAB40700.1; JOINED.
DR EMBL, U72940; AAB40700.1; JOINED.
DR EMBL, U72941; AAB40700.1; JOINED.
DR EMBL, U72942; AAB40700.1; JOINED.
DR EMBL, U72943; AAB40700.1; JOINED.
DR EMBL, U72944; AAB40700.1; JOINED.
DR EMBL, U72945; AAB40700.1; JOINED.
DR EMBL, U72946; AAB40700.1; JOINED.
DR EMBL, U72947; AAB40700.1; JOINED.
DR EMBL, U72948; AAB40700.1; JOINED.
DR EMBL, U72949; AAB40700.1; JOINED.
DR EMBL, U72950; AAB40700.1; JOINED.
DR EMBL, U72951; AAB40700.1; JOINED.
DR EMBL, U72952; AAB40700.1; JOINED.
DR EMBL, U72953; AAB40700.1; JOINED.
DR EMBL, U72954; AAB40700.1; JOINED.
DR EMBL, U72955; AAB40700.1; JOINED.
DR EMBL, U72956; AAB40700.1; JOINED.
DR EMBL, U72957; AAB40700.1; JOINED.
DR EMBL, U72958; AAB40700.1; JOINED.
DR EMBL, U72959; AAB40700.1; JOINED.
DR EMBL, U72960; AAB40700.1; JOINED.
DR EMBL, U72961; AAB40700.1; JOINED.
DR EMBL, U72962; AAB40700.1; JOINED.
DR EMBL, U72963; AAB40700.1; JOINED.
DR EMBL, U72964; AAB40700.1; JOINED.
DR EMBL, U72965; AAB40700.1; JOINED.
DR EMBL, U72966; AAB40700.1; JOINED.
DR EMBL, U72967; AAB40700.1; JOINED.
DR EMBL, U72968; AAB40700.1; JOINED.
DR EMBL, U72969; AAB40700.1; JOINED.
DR EMBL, U72970; AAB40700.1; JOINED.
DR EMBL, U72971; AAB40700.1; JOINED.
DR EMBL, U72972; AAB40700.1; JOINED.
DR EMBL, U72973; AAB40700.1; JOINED.
DR EMBL, U72974; AAB40700.1; JOINED.
DR EMBL, U72975; AAB40700.1; JOINED.
DR EMBL, U72976; AAB40700.1; JOINED.
DR EMBL, U72977; AAB40700.1; JOINED.
DR EMBL, U72978; AAB40700.1; JOINED.
DR EMBL, U72979; AAB40700.1; JOINED.
DR EMBL, U72980; AAB40700.1; JOINED.
DR EMBL, U72981; AAB40700.1; JOINED.
DR EMBL, U72982; AAB40700.1; JOINED.
DR EMBL, U72983; AAB40700.1; JOINED.
DR EMBL, U72984; AAB40700.1; JOINED.
DR EMBL, U72985; AAB40700.1; JOINED.
DR EMBL, U72986; AAB40700.1; JOINED.
DR EMBL, U72987; AAB40700.1; JOINED.
DR EMBL, U72988; AAB40700.1; JOINED.
DR EMBL, U72989; AAB40700.1; JOINED.
DR EMBL, U72990; AAB40700.1; JOINED.
DR EMBL, U72991; AAB40700.1; JOINED.
DR EMBL, U72992; AAB40700.1; JOINED.


```

OY 1741 QADEPFDSTIIQTTFPGVAL---ALGSIAPLE-----LKKTKASVGN----- 1781
Db 1423 DEDEETKEPELVGVHRNMVYIKLPHQVDGVQFMWDCCESVKRTKKSPPSGCIIAHQMGIG 1482
OY 1782 -----LFOIIOIPKSDYDPTLKS--NRYIPYVSDRYKGTIYIMEGDSEDEKYAFM 1832
Db 1483 KTLQVVSFLHLYLLODKLDFSTALYVCPPLNTALNMNNEFEK-----WQGLKDDKLEYS 1537
OY 1833 SOTTDVTSSESEY---EELDINDIYVPGSPKYKTLIEVVLPEPSGNNTTASGRNTPSDTRN 1889
Db 1538 ELATFYKRPQERSYMLQRMQEDGVMIGYEMVRNL-----AQRNVKSRKLG 1584
OY 1890 DIQNGCI---PSSKITDNEMQNLKKEFISNMLQNPNDYPSGNSSTNTNTTISR 1945
Db 1585 EILFNKALVDPGDFVVCDE-----GHILKNEASAVSKAMNSIRSRRIILGTGP- 1633
OY 1946 HNVDNNTTMTMSRDNMEENTLLPSIHGNTLYSGEESYVNVNV--NSMNDIPINRDNVY 2003
Db 1634 -----LQNNLI-----EYHCNVNFIKENLIGSKEFRNRFIN 1665
OY 2004 SGIDLINDLSGSKPIDIYDEVLRKRENELF---GTENTKRTSTQNVAKTNSDPINH- 2058
Db 1666 ---PIONGGCADSTWVDV--RVKMKRAHILYEMLAGCVQRK---DYTALTKFLPKHEHY 1716
OY 2059 -----QLELFHKWLDLDR-----HRDMCEKWKNKEDI 2083
Db 1717 VLAVRMTSIOCKLYQYLDHLTVGVNNSGEGRGKAGAKLFODFOMLSRIWTHPWCLQILDY 1776
OY 2084 LNKLEEMNKENINNSGKTYNSDNKPSHNHVLNTDYSIOI-DMDNPKTKNEITNMDTNOD 2142
Db 1777 ISKENKGYFDEDSMDEFIASDSD-----ETSMLSSDDYTKKKKKGKKGKDDSSSS 1827
OY 2143 KSTMOTILODDLEKYN 2157
Db 1828 GSGSDNDVEYIKVMN 1842

```

Search completed: May 1, 2001, 13:08:48
 Job time: 316 sec

THIS PAGE BLANK (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2001, 13:02:07 ; Search time 53.97 Seconds

(without alignments)
2837.034 Million cell updates/sec

Title: US-09-508-967-1

Perfect score: 12100

Sequence: 1 MATSGSGGTQDEDAKHVLD.....VNKKKEFEFEYPISDIWTNI 2228

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	12065	99.7	2228	2 T14029	variant-specific s
2	5074	41.9	2197	2 B71600	variant-specific s
3	4901	40.5	3078	2 T28432	variant-specific s
4	4817	39.8	2182	2 T28634	variant-specific s
5	4711	38.9	2664	2 T28626	variant-specific s
6	4084	33.8	2647	2 T28161	hypothetical prote
7	4073	33.7	2042	2 T18399	variant-specific s
8	3989	33.0	3006	2 T28625	variant-specific s
9	3937	32.5	2924	2 T18378	variant-specific s
10	3297.5	27.3	1711	2 C71625	variant-specific s
11	3192	26.4	2706	2 T28155	variant-specific s
12	2959	24.5	1729	2 T18396	erythrocyte membra
13	2863	23.7	2212	2 T28157	erythrocyte membra
14	2668	22.0	3026	2 T28431	variant surface pr
15	2518	20.8	2135	2 T14602	variant-specific s
16	1319.5	10.9	431	2 F71600	variant-specific s
17	782	6.5	440	2 E71625	variant-specific s
18	418	5.5	1435	2 A37793	erythrocyte-bindin
19	406	3.4	248	2 C71624	variant-specific s
20	360	3.0	4550	2 T18440	hypothetical prote
21	339	2.8	1526	2 A45605	mature-parasite-in
22	329.5	2.7	3394	2 T18501	hypothetical prote
23	321.5	2.7	3523	2 T18477	hypothetical prote
24	321	2.7	2441	2 D71623	erythrocyte membra
25	318.5	2.6	2401	2 T28676	thopy protein -
26	315	2.6	1070	2 T30848	Duffy receptor - p
27	310.5	2.6	1979	2 T31622	hypothetical prote
28	303	2.5	2269	2 T28677	thopy protein -
29	301	2.5	2829	2 A42771	reticulocyte-bindi

30	300.5	2.5	3724	2 T18427	hypothetical prote
31	287	2.4	1308	2 E71622	probable membrane
32	284	2.3	6642	2 T29757	protein UNC-89 - C
33	283.5	2.3	3844	2 T18402	asparagine/asparta
34	282.5	2.3	3507	2 T34513	hypothetical prote
35	279.5	2.3	1939	2 T18372	repeat organellar
36	277.5	2.3	1658	2 S55101	hypothetical prote
37	271	2.2	2485	1 H71621	serine/threonine-s
38	269.5	2.2	1631	1 S420K1	major mezozoite su
39	267.5	2.2	1726	2 A45948	major mezozoite su
40	265.5	2.2	1639	2 S05603	major mezozoite su
41	265	2.2	2657	2 T18497	hypothetical prote
42	263.5	2.2	1640	2 A24594	probable major sur
43	262.5	2.2	1045	2 T18373	erythrocyte bindin
44	262.5	2.2	1726	1 S420G	major mezozoite su
45	262	2.2	3147	2 T18674	hypothetical prote

ALIGNMENTS

RESULT 1
T14029
variant-specific surface protein 1 - malaria parasite (Plasmodium falciparum)
N:Alternate names: erythrocyte membrane binding protein 1 (EMP1)
C:Species: Plasmodium falciparum
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jun-2000
C:Accession: T14029
J.Chen, Q.; Barragan, A.; Fernandez, V.; Sundstrom, A.; Schlichterle, M.; Sahlen, A.
J. Exp. Med. 187, 15-23, 1998
A:Title: Identification of Plasmodium falciparum erythrocyte membrane protein 1 (PfEM
A:Reference number: Z17860; MUID:98080592
A:Accession: T14029
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2228 <CHE>
A:Cross-references: EMBL:AF003473; NID:g2961467; PID:g2961468; PIDN:AAC05730.1
C:Genetics:
A>Note: FCR3S1.2-vari

Query Match	99.7%	Score 12065;	DB 2;	Length 2228;
Best Local Similarity	99.8%	Pred. No. 0;	Mismatches	5; Incls 0; Gaps 0;
Matches 2223; Conservative	0;			
QY	1	MATSGSGGTQDEDAKHVLD	EFQKYHDEVHGEAKNYVELKGSLSLA	SILGETAFYVKS 60
DB	1	MATSGSGGTQDEDAKHVLD	EFQKYHDEVHGEAKNYVELKGSLSLSLA	SILGETAFYVKS 60
QY	61	MOTESKYTELIEANSKRNECKKDGKNDVRSVKEQAGYDNKKMKK;	INGMTCAFFRLH 120	
DB	61	MOTESKYTELIEANSKRNECKKDGKNDVRSVKEQAGYDNKKMKK;	INGMTCAFFRLH 120	
QY	121	LNKNFPPNNNSDSSKAKHDLAEVCM	AKYEGESIKTHYPYDSKYVGSDFPCKTMLAR 180	
DB	121	LNKNFPPNNNSDSSKAKHDLAEVCM	AKYEGESIKTHYPYDSKYVGSDFPCKTMLAR 180	
QY	181	SPADIGDIRGDLVYLNKKK	KKONGKETREKLDEKLKEIFPKIH	DN.KDKAOKRYNND 240
DB	181	SPADIGDIRGDLVYLNKKK	KKONGKETREKLDEKLKEIFPKIH	DN.KDKAOKRYNND 240
QY	241	EDPNFYKLEDMWTANRET	VGAMTCSKELDNSSYFRATCNDTGGP	;QTNHKKCRCDK 300
DB	241	EDPNFYKLEDMWTANRET	VGAMTCSKELDNSSYFRATCNDTGGP	;QTNHKKCRCDK 300
QY	301	GANACKPRAGDGDVITVPT	FDYVPOYLPMFEEMADEPCRRKKKLE	TEKCCRRKDS 360
DB	301	GANACKPRAGDGDVITVPT	FDYVPOYLPMFEEMADEPCRRKKKLE	TEKCCRRKDS 360
QY	361	ERYYSRNGYDEQFTSRK	QVYRMKGKGTDFACGSYENWIDNOR	KJFDOKKYYTEIS 420
DB	361	ERYYSRNGYDEQFTSRK	QVYRMKGKGTDFACGSYENWIDNOR	KJFDOKKYYTEIS 420

Qy 421 DGGRRKRAVGGTTKYEGEKSFEYKLEKNDGYGVDAFLGLNNNEKACKDITDGGKINFK 480
Dy 421 DGGRRKRAVGGTTKYEGEKSFEYKLEKNDGYGVDAFLGLNNNEKACKDITDGGKINFK 480
Qy 481 EVNNGGCVVGGSGTSGASSTNDENKCTFYRSEYCPQPCPGCYQHKGNOMERKTAKK 540
Dy 481 EVNNGGCVVGGSGTSGASSTNDENKCTFYRSEYCPQPCPGCYQHKGNOMERKTAKK 540
Qy 541 MRMSKLVKPIGKMWLLSLKLVKVMWMLIKNNKEFCITLONSSDGSVGVYTTGASGN 600
Dy 541 MRMSKLVKPIGKMWLLSLKLVKVMWMLIKNNKEFCITLONSSDGSVGVYTTGASGN 600
Qy 601 SEKKELYDEMKCYKHNEVQKVNVOGEVEDEDELKAGAGLICILPPKKNKEVSEAKSQNN 660
Dy 601 SEKKELYDEMKCYKHNEVQKVNVOGEVEDEDELKAGAGLICILPPKKNKEVSEAKSQNN 660
Qy 661 HADQKTFHDFYVVAHMLDSTHMKTRKLSCTSDGKTAKKCRNGCKKCDCEKVKYQ 720
Dy 661 HADQKTFHDFYVVAHMLDSTHMKTRKLSCTSDGKTAKKCRNGCKKCDCEKVKYQ 720
Qy 721 KETEMKPIKDHFTQEGIPEGYFTTLELILKLOFLKEDTEENTENSLDAEAEFLKHLQ 780
Dy 721 KETEMKPIKDHFTQEGIPEGYFTTLELILKLOFLKEDTEENTENSLDAEAEFLKHLQ 780
Qy 781 KILKLENNNLAVVNAAGTEQKTLMDKLNHELNDATKCDCLPEEDSKSRGSADPSPI 840
Dy 781 KILKLENNNLAVVNAAGTEQKTLMDKLNHELNDATKCDCLPEEDSKSRGSADPSPI 840
Qy 841 FIPRPEEKEDENEDDEDEYRDEBETAKETEGSATDTTSLDVCPLVGVVLTKNESL 900
Dy 841 FIPRPEEKEDENEDDEDEYRDEBETAKETEGSATDTTSLDVCPLVGVVLTKNESL 900
Qy 901 QDASLKYGNNRSLGMRCVPSGEPTSSDKNCAICVPRRRRLYIKKIYDMATKTESP 960
Dy 901 QDASLKYGNNRSLGMRCVPSGEPTSSDKNCAICVPRRRRLYIKKIYDMATKTESP 960
Qy 961 QASGEASSTIGSTTPDPSKALLKAFVESAAIETFLMHRYYKEEKKAQAEGAGHLPR 1020
Dy 961 QASGEASSTIGSTTPDPSKALLKAFVESAAIETFLMHRYYKEEKKAQAEGAGHLPR 1020
Qy 1021 VEGSPEDPDDELKEGKIPGGLRQMYTTLGQYRDLIFSSSNTOTYSKKTSPSSNLT 1080
Dy 1021 VEGSPEDPDDELKEGKIPGGLRQMYTTLGQYRDLIFSSSNTOTYSKKTSPSSNLT 1080
Qy 1081 KNIYLLASGSTEOERREKNNKYEIKNFRKCTERSAPNLVHPOTWMENNGKIYHGWVC 1140
Dy 1081 KNIYLLASGSTEOERREKNNKYEIKNFRKCTERSAPNLVHPOTWMENNGKIYHGWVC 1140
Qy 1141 ALTSKDKIAKGVERRPKKIENPENLMDANKRPPOYQYTNVNLKDENSGTSPTTQTA 1200
Dy 1141 ALTSKDKIAKGVERRPKKIENPENLMDANKRPPOYQYTNVNLKDENSGTSPTTQTA 1200
Qy 1201 SSDNTPTTLTHFVFRPYFRFEEEMGESFCBERKKRLQIYDCKVENGDVGRCSGDEA 1260
Dy 1201 SSDNTPTTLTHFVFRPYFRFEEEMGESFCBERKKRLQIYDCKVENGDVGRCSGDEA 1260
Qy 1261 CDSISTHDYSTVSPFNCPCGCKHCSYRKAWEKKIEFHKSNAVGOOKTATATNNNGTF 1320
Dy 1261 CDSISTHDYSTVSPFNCPCGCKHCSYRKAWEKKIEFHKSNAVGOOKTATATNNNGTF 1320
Qy 1321 DKEFCCTLTPDAKFLERLKNGPCYTNKEYGDDIDFEKDSCTFOHTEYCGCPKPKYT 1380
Dy 1321 DKEFCCTLTPDAKFLERLKNGPCYTNKEYGDDIDFEKDSCTFOHTEYCGCPKPKYT 1380
Qy 1381 NCONGNCVSLGNCNGDGSIDAKETAKMRSSTTDVVMRVSDNDTTFEGDDLKDAQOH 1440
Dy 1381 NCONGNCVSLGNCNGDGSIDAKETAKMRSSTTDVVMRVSDNDTTFEGDDLKDAQOH 1440
Qy 1441 ANIFKGRKDWKCGYCVGVICBOTNINERTDKEYIOIALKRRVNTLEBYNKTND 1500
Dy 1441 ANIFKGRKDWKCGYCVGVICBOTNINERTDKEYIOIALKRRVNTLEBYNKTND 1500
Qy 1501 KISHCIKKGESKINCCEKNSKLEKWEIKKIAEMENIKKRFNDQYENKQDPYNYKSI 1560

Db 1501 KISHCIKKGESKINCCEKNSKLEKWEIKKIAEMENIKKRFNDQYENKQDPYNYKSI 1560
Qy 1561 LEEILPKIAVVDNDNYIKLGVFENSGCTLISNTONNKENDALDCHLKRIGVAKKCPG 1620
Dy 1561 LEEILPKIAVVDNDNYIKLGVFENSGCTLISNTONNKENDALDCHLKRIGVAKKCPG 1620
Qy 1621 KPSEKOSDCKEPPPLPDEEDONPEENTLEBPCKCPPTTOPPEKGETCGNKEEKDEK 1680
Dy 1621 KPSEKOSDCKEPPPLPDEEDONPEENTLEBPCKCPPTTOPPEKGETCGNKEEKDEK 1680
Qy 1681 KESEEPAKESGPAEPAEPAESETETNFPPEPGTGAAPPSTPAAPPPTPPPLRP 1740
Dy 1681 KESEEPAKESGPAEPAEPAESETETNFPPEPGTGAAPPSTPAAPPPTPPPLRP 1740
Qy 1741 QADEPFSTIIQTTPRGVALALGSIAFLFKKTKASVGLRQILQIPKSDYDIPPLKS 1800
Dy 1741 QADEPFSTIIQTTPRGVALALGSIAFLFKKTKASVGLRQILQIPKSDYDIPPLKS 1800
Qy 1801 SNRYIPIVSDRYKGGTYIYMEGSDDEDKAFMSDPTDVTSSSEYEELIDINDIYVPSPK 1860
Dy 1801 SNRYIPIVSDRYKGGTYIYMEGSDDEDKAFMSDPTDVTSSSEYEELIDINDIYVPSPK 1860
Qy 1861 YKTLIEVVLPEPSGNNTTASGKNTPSDTRNDIQNDGIPSSKITDNEWNLKKEFTSNMLQ 1920
Dy 1861 YKTLIEVVLPEPSGNNTTASGKNTPSDTRNDIQNDGIPSSKITDNEWNLKKEFTSNMLQ 1920
Qy 1921 QPNVNDYTGNSSTNTNTTTSRANVDNNTNTTMSDNNNEULLPSTHDGLVYGE 1980
Dy 1921 QPNVNDYTGNSSTNTNTTTSRANVDNNTNTTMSDNNNEULLPSTHDGLVYGE 1980
Qy 1981 YSYNVNMYNSMNDPIRNDNNVYSGIDLINDLSGKRPIDYDVLARKENELGTEGNTK 2040
Dy 1981 YSYNVNMYNSMNDPIRNDNNVYSGIDLINDLSGKRPIDYDVLARKENELGTEGNTK 2040
Qy 2041 RTSTQNAKTTNSDPIHQLELFRKWLDRHRDCEKWKNEKEDILINKKEENKENINNSG 2100
Dy 2041 RTSTQNAKTTNSDPIHQLELFRKWLDRHRDCEKWKNEKEDILINKKEENKENINNSG 2100
Qy 2101 KTYNSDNKPSHNHVLNTDVSQIDMDNPKTKNETTNDNDKSTMTDILDDLEKYNDPY 2160
Dy 2101 KTYNSDNKPSHNHVLNTDVSQIDMDNPKTKNETTNDNDKSTMTDILDDLEKYNDPY 2160
Qy 2161 YYDFEEDDIYHDVDVEKSSMDDIYVDHNNVTSNMNOVPTFMHIEHNNVNNKKELFEEY 2220
Dy 2161 YYDFEEDDIYHDVDVEKSSMDDIYVDHNNVTSNMNOVPTFMHIEHNNVNNKKELFEEY 2220
Qy 2221 PISDIWNI 2228
Dy 2221 PISDIWNI 2228
RESULT 2
B71600
variant-specific surface protein 1 homolog PFB1055c - malaria parasite (Plasmodium fa
N:Alternate names: erythrocyte membrane binding protein 1 (EMP1)
C:Species: Plasmodium falciparum
C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
A:Accession: B71600
R:Gardner, M.J.; Tettein, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.
; Pereira, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H
Science 282, 1126-1132, 1998
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600; MUID:99021743
A:Accession: B71600
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-2197 <GAR>
A:Cross-references: GB:AE001434; GB:AE001362; MID:g3845341; PIDN:AAC71996.1; PID:g384
A:Experimental source: Clone 307
C:Genetics:
A:Gene: PFB1055c

1932 KDVENDKSGGIPENTQ-----DNTLYFDKPEEKPEITSHIDNLLNGEYS 1978
1983 YNVM-NVSMNDPILNRNNYSGDILNDSGKPRIDYDEVLAKRENELFCGENTKR 2041
1979 YNVMKSNMSMD-PRYSNNYSGDILNDSGKPRIDYDEVLAKRENELFCGENTKR 2037
2042 TSTONVAKTNSDPINHOLEFHKMLDRHRDMCKEKKKEDILNKLKEEMKENINSNGK 2101
2038 TSHSVAKNTSDPILNQLNLFHTWLDHRDMCKEKKKEDILNKLKEEMKENINSNGK 2095
2102 TYNSDNKPSNNHNTVDSIOIDMDNPKETKNETNMOTNODKSTMDITLDDLEK-YNDPY 2160
2096 THPSDS-----NKTINTDVSIOIHMDNPKETKNETNMOTNODKSTMDITLDDLEK-YNDPY 2142
2161 YDDYEDDIYHDYDVEKSSMDIYVDH--NNVTSNNMDVPTKMHIEKNINNNKEIFEE 2218
2143 YDMDYDD-YYDVA-----DHDISTVDTNADVPKVOIEKD-VNTR--LVKE 2187
2219 EYPISDIMNI 2228
2188 KYPIADYWDI 2197

RESULT 3
T28432
variant-specific surface protein 1 - malaria parasite (Plasmodium falciparum)
N:Alternate names: erythrocyte membrane binding protein 1 (EMP1)
C:Species: Plasmodium falciparum
C>Date: 15-Oct-1999 #sequence, revision 15-Oct-1999 #text-change 09-Jun-2000
C:Accession: T28432
R:Su, X.Z.; Heatwole, V.M.; Wertheimer, S.P.; Guinet, F.; Herrfeldt, J.A.; Peterson, D.S.
Cell 82, 89-100, 1995
A:Title: The large diverse gene family var encodes proteins involved in cytoadherence and
A:Reference number: Z20487; MUID:95330813
A:Accession: T28432
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3078 <SUX>
A:Cross-references: EMBL:LA0608; NID:9886374; PID:9886375; PIDN:AA75396.1
C:Genetics: var-1
A:Gene: var-1
A:Introns: 2611/3

Query Match 40.5% Score 4901; DB 2; Length 3078;
Best Local Similarity 36.6%; Pred. No. 1.2e-214;
Matches 1174; Conservative 277; Mismatches 631; Indels 1128; Gaps 83;

QY 4 SGGSGTQDEDAKHLVDFGQKVHDEVHGEAKNVSELSGLSLASLIGETAFTVKSMTQ 63
12 AAGGDDIEDESAKHMFPRIGKDYDKVEAKERKGGQRLS-----EAKFEKNESDP 65
QY 64 ES-----KYTELEANSKNPCK-----KDGKNDVDRFSYKEQAGYDNKKMKC 107
66 QTPEDPCDLHKYHTNTTNN-VINPCADRSVRSDEYGCSTHNRKISIQGQDNKG-- 121
QY 108 SNGTCAPFRHLHCNKNFPMNSNDSSKAKHDLAEVCAAKKEGSIKTHYKPYDSKY 167
122 ---ACAPYRRLHVCNDLNEQIEPIKITNT-HNLLVDCMAAKFEGOSITODYPRKYATY 176
168 PGSDPFCPLMARFADIGIIRGRDYLGNKKKKKQNGKTEREKLEOKLEIFKFIHDN 227
177 GDSQSQCMTMARSFADIGIIRGRDYLGNKKKKKQNGKTEREKLEOKLEIFKFIHDN 231
QY 228 LKDKKQKRYNGDEDPNFYKLRDMWTANRETVGAMTCSKEIDNSSYFRATCNDTQGP 287
232 LKNGAE--RYG--NDPEFKLRDMWTANRETVGAMTCSKEIDNSSYFRATCNDTQGP 281
QY 288 SGNHNRCKDKGANKGKRAAGDVTIPTYFDYVPOYLRFWEBAEDFCRKKKKLE 347
282 ERTKGYCRKNDQ-----VPTYFDYVPOYLRFWEBAEDFCRKKKKIK 325

QY 348 NLEKQCRGKDKSDEYRYCSRNGYDCEQTSIRKKGVRMKGCTDCEFAAGSYENNIDNRK 407
326 DVKRNCRGKDKEDDRYCSRNGYDCEQTSIRKKGVRMKGCTDCEFAAGSYENNIDNRK 385
QY 408 QPDKQ-KYTYETI-----SDGGKRRKAVG--TTKEGYEKSFYEKLKDGCTYDAF 458
386 QPDKQKRYDEIRKYEENGAGSGGSRKRDAGTITTYNDGYEKKEFYDELNKSERYTDKF 445
QY 459 LGLNNEKACDITD--GKIFNFEVNSGGVVGSGGSGTSGASTDENKGFYRSRYC 516
446 LEKLSNEETCKVNDDESGTIDFKNV-----SDTSGAGSINVESTGTYRSKYC 496
QY 517 QPDCDQVQ--HKG--NOMERKTRVKKRWSKLYK--INGKAVLLKSLKLVKQDM 569
497 QPDCDQVKKVNGSGSSNEMEEKNN-GKCKSGKLYEPKPEKGTITILKSGKHND-- 552
QY 570 LKKMKKEFLTONSSDGSVSVTTGASGSGSEKELDYDEKCYKHNQVQVNOGEYEE 629
553 IEKLNKFCDEKNGDTJNSGSGTGGSGSGSGRQELYEEKCYKGBDVVYVGHDEDEE 612
QY 630 DDELKGAAGLCTLPNPKKNEVEAKSONNHADIKTFHDFYVVAHMLKDSIHMRTK 689
613 DYENKNAAGLCTLPNPKKNEVEAKSONNHADIKTFHDFYVVAHMLKDSIHMRTK 671
QY 690 RLKSCISDQKTMK--RNGCNKRCDFEKKVYKQETEMKPRIDHFKTQEGIPGYFTTLE 748
672 KLQCLQNGNRITKCGNNKCNNDSCFKWITQKDEMGKIYQHKRQIKNKSGSDMTAE 731
QY 749 LI-----LKQFLKED-----TEENTENSIDAEAEELKHLKILKLENNILAV 793
732 LIPDHVDVLYQNLQEEFLKDSSEDAESEKSENSIDAEAEELKHLKILKLENNILAV 791
QY 794 VNAG--TEOKTMDKLNHLNEDATKCKDCPLREBDD----- 828
792 VGGVTEOKTMDKLNHLNEDATKCKDCPLREBDD----- 828
QY 829 ----- 828
852 SGNKRYPLANKVAYONHNKAKTQLASRAGSALRGDISLAQFKNGRNGSTLKGQICKIN 911
QY 829 -----SKRNSADPS-----PDIFIP----- 843
912 ENYSNDSKNGSGPCGTGDDGHDGVRMRIGTGWENIEGKQTSYKNVFLPRRHEMCTSN 971
QY 844 ----- 843
972 LENIDVGSVTKNDKASHSLDVLQDLAKTDAEILIKRYKQDNNTIQLDPIQKQDEAMCR 1031
QY 844 ----- 843
1032 AVRYSPADLDGIIRGRDMDDEKSTDMETRLITVPFKNIKEKHGINDNPKYTGDESKP 1091
QY 844 ----- 843
1092 AYKRLRADWMEANRHOVRAWKCATGIIICPMFVDYIPQRLKMTWMAEWYKASOE 1151
QY 844 ----- 843
1152 YDKLKKICADCMKSGDKCTQGDVDCGKCAACDKYEEIEKMEQWRKISDKYNLLYLQ 1211
QY 844 ----- 843
1212 AKTSTNPGRTVLDDDDPDTQDMVDFLPIHKASIAARVLVYKRAAGSPTEIAAAAPTTPY 1271
QY 844 ----- 843
1272 STAGYIHOEIGYGOEOTOFCEKKHGASTSTTKENKEYTEFKORPREYATACDINRS 1331
QY 844 ---RPEEKD----- 850
1332 QTEPFRKKEENVESACKIVEKIEGKNGRTTVGECNPKESYPMDCKNNDISIDGACMP 1391
QY 851 -----DEN-----EDDEDEEVRD-- 863

Db	1392	PRROKLCUYIAHSESOTENIKTODNLKDAFIKIAAAETLSMOYIKSKNDSEAKILDLGL	1451
Qy	864	-----DEBTARE-----TTEGS-----	875
Db	1452	IPSOFLRSMMYTFEDYRDICLNTDISKQNDVAKAKDKIGKFFSKDGSKSPGSLRQEW	1511
Qy	876	-----ANDT-----	879
Db	1512	KTNGPEIWKGMCLATKYVYDTONKRRIKINDYSYDKVNOSONGNDPSEFAKPOFLRM	1571
Qy	880	-----	879
Db	1572	IEWGEFCAEROKKENI1KQACNEINSTOOCNDAKHCNOACRAYOEVENKKFEFSQT	1631
Qy	880	-----	879
Db	1632	NNFLKANNVQPODEYKGYEKKGVOPIQONEVLLOKCNMKSCMDGAVLSVSKPEP	1691
Qy	880	-----TSLDVCPIYKVLTKDNES1QDA	903
Db	1692	GKYAHKYPEKDCYOGKHVPSIPPPPPVQPOPEAPETIYVDCSYI-KTLFQDNTNPFDA	1750
Qy	904	CSLKYGGNNSLGRCY---TPSGEPT---SSDRKALCVPPRRRLYIKIYVAKRT	957
Db	1751	GLGY-KGTAASSWKCIIPSDTKSGAGATGKSSDGSISICIPRRRLRYLKGLOEMATAL	1809
Qy	958	ESPOASGEASSTGSTTPPDSKEALLKAFVESAL1ETFEFLMHRYKEEKA0AAGAGHG	1017
Db	1810	--POGEAAPHSHA-----DOLRNAFIQAAL1ETFLDRKKEKKPGC-DGSOOA	1858
Qy	1018	LPRVEE-GSPEDPEDK-LKEGRIPDGLKROFYLGDYRDLFSGSNDTYSVKDTPS	1074
Db	1859	LSQLTSTYSDEEDPDPKLLQNKIPDPFLRLMFLYTLGDYRDLVHGN--TSDGNTG	1916
Qy	1075	SSNDLKNYVLASGSTEOREKMNKKEK--KNFKCSTESAPNLVSHPOTMENNGK	1133
Db	1917	SNNN--NIVLEASGNEDMOKIOEKIEOILPKNGGTPIVPKSS--AOTPDKMNNEHAE	1970
Qy	1133	YIMHGMVCALT---SKDKLAKGVEKKPOKIEKPEMLMDE-----ANKK	1175
Db	1971	SIMKGMICALITYIEKNPDTSARQDE--NKIEODEYEKFFEGSTADKIGTASTPTGYK	2027
Qy	1176	POYOYTNVKLDENSGTSPRTTOTOASSDNTPTLTHFVRPRTYFRWFEEMGSEFCRERKK	1235
Db	2028	TOYOYEKVKLEDDISG-----AKTPSASDTP-LDSOFVLRPFRLEMGONFCKKXKH	2081
Qy	1236	RLKQIKYDKVEKNDVG-----RCSGDEACDSISTHDIYVPSFENCPCGCHGCSY	1287
Db	2082	KLAQIKHECKYEENGSGSRGGITROYSGDGEGACNMLPKNGTVADLEKPCACAPCSSY	2141
Qy	1288	RKWIERRKIEBHOSNMYGOOKTDAPRNNGNTFDRKFCFTLETFWPAKAFELRLKNGSK	1347
Db	2142	RKWLIESKGEKEKEKERYEBOK--DKCVNSNKHNDGFCFELTTSKAKOFLTL--GPOK	2198
Qy	1348	TNKEYGGDDIDFEKDSKTFPOHTEYCGCPKFTKCNQNGKCGVSLGNGNDGKSDIAKEI	1407
Db	2199	PNNVEGKTIP---DOKTFKHTHDCDOPCLKFSYCNKKKDE--NSKGTDCRKNKNSIDARDI	2254
Qy	1408	AKMBSSTTDVVMRVSDNTJTFBGDDLKDAQOHANTFKGIRKDWKCGYGVGDICEOTN	1467
Db	2255	ENGVDSTYV-LEMRVSAJSKSGFMNDGJLENAACRAGIFEGIRKDEKMCRWVCYVVCPEPN	2311
Qy	1468	INERTRDKEYIQTALRKRNVENPLEBYNKNINDXISHC1KKGGSKCIKNGSKNKKCLEK	1527
Db	2314	VNGEAKGH11Q1ALVKRWVEYFEEDYENIKIKRISHRIKINBEISPT1---KN--CYEK	2367
Qy	1528	WIEKIAEWEIKRRFYNDQYENKDDOPYNKS1LEELIKPIAVNPODVAIKLAFVENS	1587
Db	2368	WVDQKREKMKITIERFQDKKNDSSDDNVASTLELILQIYIDANAANKNVIYLSFGVNSC	2422
Qy	1588	GCTLISNTQ--NKNKENDADCMKLGLGVKAKNCPG--PSEKGEOSCKEPPLEDE--	1639

Db 2428 GCGSAAANQNKNGEYKDAIDCMKLLKXKIDCECKHHQTSDTCSDFI QOQTLDEDTLD 2487

Qy 1640 ---EDQNPBENTLEPPKFCPT--TQPEEKGETCGNKEEKDCKEI:SEPAKEESGP 1694

Db 2488 DQIEFEAKKKMM--PKICEVNLKTAQODEGG--C-----VPAHISEEPAATSGK 2535

Qy 1695 AAEEAPPAESETETNPPPPGTCAPAPRPTAP--PTDTPPPLRIQADEPFDSTIL 1751

Db 2536 ETPEDTPALKEPEEAVPPEPP-----PQEKAPAPIQPQPTPTPTQLIDNHNVTAL 2599

Qy 1752 QT-TIPFGVALAGSIAFLFKKTKKASVGNLFOLQIPKSDYDLPPL:SNRRIPYVSD 1810

Db 2590 VTSTLAMSVCIEFATFTFYLLKTKKSSVGNLFOLQIPKSDYDLPPL:SNRRIPYVSD 2649

Qy 1811 RYKGTATYMEGSDSDDEDYAFMSDTDYTSSESEYEELINDIYVPGSKYKYLEVLE 1870

Db 2650 KRGKRYIYLESDSTDS-GYTDHYSDITSESEYEEMDINDIYVPGSKYKYLEVLE 2708

Qy 1871 P-----SGNNTTASGKNTPSPDRNDIQNDGIPSSKITDINQOLKEETISN 1916

Db 2709 PGGNNTTASGKNNTTASGKNTPSPDRNDIQNDGIPSSKITDINQOLKEETISQ 2768

Qy 1917 MIONPNVNDYTSGNSNSTNTITTSRHVNDNTNTSRDMMNEEN.LJPSIHDLGX 1976

Db 2769 YLQSEPNTEPN-----MLGYNDNTNHTPTSHHNVKEPFTMSIHDLRLF 2813

Qy 1977 SGEESYVY-----NMVN-----SMNDPIRDNVNVSGIDLINDSLSGKPI 2019

Db 2814 SGEENYNDYEMGNNPINISDSTNSMDLSTNSNHPYDKNDLVSGLIDLINDALSQNH-I 2872

Qy 2020 DIYDEVLARKENELFGE-NTKRTSTQNVAKTNTSPDPIHNOLEFHKWLDHRDMCEKWK 2078

Db 2873 DIYDMLKRRKELEFGTKHHTFTYVNAKPARDDPTTQNLNFKHKLDRHRDMCEKWK 2932

Qy 2079 NKEDILNKLKEMKNENINNSGKTYNSDNKSKSHNVLVLTQVSIQIDNMNPKTKNEITMD 2138

Db 2933 NNHERLPKIKELM--ENETHSGDI--NSGIPSGNHLVLTQVSIQIDNMNPKTKNEITMD 2988

Qy 2139 TNQDSTMDTILDDLEKYNPPYDFEDDIYHVDYKSSMDIYDHNNTSNMNDV 2198

Db 2988 TNPDKSTMDTILDDLEKNEPYYDFEDDIYHVDYKSSMDIYDHNNTSNMNDV 3048

Qy 2199 PTKMHIEHNVNKKKEIFEEXYPISDIWINI 2228

Db 3049 PTKMHIEHNVNKKKEIFEEXYPISDIWINI 3078

RESULT 4

128634

variant-specific surface protein 7 - malaria parasite (Plasmodium falciparum)

C.Species: Plasmodium falciparum

C.Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000

C.Accession: T28634

R:Su, X.Z.; Heetwole, V.M.; Wertheimer, S.P.; Guinet, F.; Herrfeldt, J.A.; Peterson, Cell 82, 89-100, 1995

A>Title: The large diverse gene family var encodes proteins involved in cytoadherence

A.Reference number: Z20487; MUID:95330813

A.Accession: T28634

A.Status: preliminary; translated from GB/EMBL/DBJ

A.Molecule type: DNA

A.Residue: 1-2182 <SD>

A.Cross-references: EMBL:L42636; NID:g886379; PID:g886380; PIDN:AAA75399.1

C.Genetics:

A.Note: var-7

Query Match 39.8%; Score 4817; DB 2; Length 2182;

Best Local Similarity 45.9%; Pred. No. 4,86-211;

Matches 1096; Conservative 298; Mismatches 607; Indels 388; Gaps 93;

Qy 4 SGGSG-----GTQD-----EDAKHYLDFGQKVDH--VHGAKNRY:ELKGSLSLAS-I 50

Db 11 SGGSSSGGKKDTSEYIIVVSDAKDLIDRVGKRYVEEKYKNGDARKYI:ALKGNTANGR 70

Qy	51	LGEAFYVKSQTESKYTELLEANSKRNPCKDKGNDYDRS-----VKEOAG	99
Db	71	SSEASTSIETCTLVEKEYERYENGDBGRHPCKRDAKNEDYRPSDTLGCGTYNRKIDSOO	130
Qy	100	YDNKMKCSNMCTAPPRRLHLCKNPNMNSDSSKAHODLAEVMAAKKEGSIETH	159
Db	131	GDNK-----GACAPYRRLHLDY--NLESIDTSTYHKLLEECVMAKKEGSIETH	181
Qy	160	YPKYDSKPSDPMCMMLARSPADIDDIRGHDILG--NNKKONGKETRERLEOKL	217
Db	182	YTHQRTHMEDASOLCTYLARSPADIODIYRGDYLGLGDNKEKO-----RKLLEOKL	235
Qy	218	KEIEFKIH-DMLKDKEAOKRYNED-EDPNFYKLREDMWTANRETYWAGTSCKELDSSY	275
Db	236	KDIEFKIHKDVMKNGAERYERIDAKGDFOLREDMWMTSNRETYWALICHAPEAWFY	295
Qy	276	FRACNDTGOOPSOITHKCKCDKAGNACKPAGGCVDTYPTYPDYDYPOLRMEEMA	335
Db	296	IKTAGN-GKG-----TNGOCHC-----TGGD-----VPYFDYDYPOLRMEEMA	335
Qy	336	EDFCRKKKKKLENEKCKCRKDKSDERYCSRNGYCEOTISRKGVVRMGKCTCPAC	395
Db	336	EDFCRKKKKKLENEKCKCRDYEDN--LYCSNGYCTCTYTKKGLYGBECTCSWC	392
Qy	396	GSYENWIDNORKRODKOK-KYTRKESDGG-----RKRRAVGATTKYE--GYEKSFEKL	447
Db	393	RMYETWIDNCKEPLKOKRYETETISGSGSKSPKRTKPAANSSSSDNGYESKFFKL	452
Qy	448	KNDYGYVDAFLGLLNNEKACKDITDGGKINFEVNSGGGVGGSGGTSGASGTNDENK	507
Db	453	KEYGVODDKFLILNEGICOKOPOVG--NEKADN-----VDFTNEYV	495
Qy	508	GTFYREYCOCPRCQVOHNGNOMERKTYVKRMRSKLYKPLNGMYLLKSLVAVDM	567
Db	496	KTFSTELCEPCPCPGIE-KGPRW-KYAGDKTCSAATKYDPDNITDITVLPDSQ	552
Qy	568	MILKNMKEFLTONSSDSGVSVYTTGASGSGSEKKELYDEMKCYKHNENOVAVOGEV	627
Db	553	QNLKRYNFC-----EKGAGGGQIHK-----WQCY-----	580
Qy	628	EBDDDELKGAGGLCTLPNPKKNREVSFAK---SONNHADIKTEHDFEYVVAHMLKDS	683
Db	581	-----DEHR-----PSSKNNNVCVEGTWKPQCKOT--VAKSYVFEWMDVHMLDS	626
Qy	684	IHWRTKRLKSCI---SOGKTMKCRNCKKCDCEFEKVMVOKETEMPKYIDKHKKOEGI--	738
Db	627	VEKITE-LSKTINNTNGTCRNNNCKCTKDCGFOKVAERKQOEWMALKDHFCKOTDYO	685
Qy	739	-----PEGYFTYTELILK-----LOFLKDETEENTENSIDAEAEELKHLQIKLE	786
Db	686	OKGLIVSPYG---VLDYLVKGNLLQNTK-DVHDT-----DDIKHKKL--LD	729
Qy	787	NENNLAVVNACTEOKTLMDKLNLHELNDATCK---DCPLREEDKSNGRSADPSPIFI	842
Db	730	EDEAVAVVLGCKDWTIT-DKLLOHEKEOAEQCKOQOECEKRAQOESRSASRETERDT	788
Qy	843	PRP-----EKEEDENMEDDEVERDEDETAETEGSATOTITSLOV-----	885
Db	789	QOPADSAGEVEEEDDDDDYDEDEDDDDVOEEBEKE--EGTVLETEVTEVEETVEQ	846
Qy	886	-----CPYVGVKLVKDNESLQDACSUKY--GGNNSRLGRCVTPSGEPTSSDKNGAICV	938
Db	847	EGVAPCDIYVG-LFEDDKSLKAEACGLKYGCGKEKPPMKCVTPSGVSTATSGRGAICV	905
Qy	939	PPRRRLYVGLISQMAR-----GGDETTEVSEATSPASSESSEKLTAPIESALET	959
Db	906	PPRRRLYVGLISQMAR-----GGDETTEVSEATSPASSESSEKLTAPIESALET	959
Qy	996	FFLHNRKKEEKKAVAL-OBGAGHCLPRVBEESPEYDEEDLKE-GKLPDGLRQMYTTIGD	1053
Db	960	FFLHNRKKEEKKPPATDGDGAGLVS-LPEBSPBEDDQOTLOOTGVIPDFLQOMMYTTIAD	1019

QY	1054	YRDLFFSGSNNTT-SVKOKTPSSNDMLKATVLLAGSTFQOREKMN----	КХКЕІКНР	1108
Db	1020	YKDLIYSGSSNDTSPSTTKQTPSSSNDMLKATVLLAGSTFQOREKMNQIOAKIKILGA		1079
QY	1109	KCSTERSAPNLVSHR-OTWMENNGKTYIWMGVALCSKQIKQYGEK-PQKIEPNEMJL		1166
Db	1080	TSQVPPTKSNVSKPPOOTWMENLAKIDIMANMVCALITYKENDARGISAKIEQKDKLKMJ		11399
QY	1167	DEANKPKRPOYQYATNVKLJENSGTSPRTQTOASSDNT-PTTLTHPYKRPTEYRMEE		12244
Db	1140	DEANKNPTEIKYQYATNVKLJEDSG-----AKSNDTQPTPLKNFVEIPTFRMLHE		1190
QY	1225	WGESCFRRKRRKQIYVDOCKVENGDYRGCSGDGEACDSISTHDYSTYPSFNCPECCGHC	1284	
Db	1191	WGNSECFERRAKRQIOIHCECDEJGE-KOYSGGEYCEBEIFSKOYNVLQDLS-SSCAMP	1248	
QY	1285	SSYRWIERKKTEPHKOSNMYGOOKTD-----ATNNOCNTDKECKTLETPWPA	13355	
Db	1249	RLYKWTIEKKTEYERKQOKAYEOKSAYENEOKQOCOTOSNNA--NEBSRLGASPTAA	1306	
QY	1336	KFELELKNGPCKTKMEY-GGDD-IDFEKDSKTFPHTEYEGCGPKFTNCOQNGVSGJL	13922	
Db	1307	BELOKL-GSCKNDNGVENEDNKIDPKNDKFRKAHSCDPCPIYGVKQONGHC-VESA	1363	
QY	1393	NG-NCDGKSIDAKEIAKMRSSTTDVVMYVSDNTTFE--GDCLKDAQHANIFKIRK	1449	
Db	1364	NGKECKNNK-ITAEIDIKNKTDPMNENIMYVSDSTWTFELGD-----CKSGSIEGIRK	1417	
QY	1450	DVMCGVYCGVDIC-QOTMINERTGKEYIOJRALFRWVENPLFEDYKNIDKISHCIR	1507	
Db	1418	DEMKCANCGVDITLTKRKIKNOEGDKITIMKELLKMWLEFLIEDYNRIRKIKILCTK	1477	
QY	1508	KGECSKJNGCEKNSKLEKMIKIEKLAEMENIKRPNDOYENKDDOPYNVASILEELIP	1567	
Db	1478	KEDGCKCLG-----CIEKHWQEKIKEMKINDYTLBEDYKND-----GNLTNLNLEQ	1525	
QY	1568	IAYVNDODNVKILC---VEENSKGCTLSJNQNKENDADIDMLKLGVKAKNCPGKPS	1623	
Db	1526	FQYRTPEKNAKPCDGDIDOKFTSGJLSTNSQGNNDLVCLLTKLOKRISECKEQHS	1585	
QY	1624	GKQSDC-----KEP---PRLPDEDDQPRENTELEPKFCPTTQPREEKGGE--TGC	1671	
Db	1586	GOTQPCPNSSLSGKESTLVEDVDYDEONP-ENKYEOKFCPDMKEPKKEDDEVGTCG	1644	
QY	1672	NKEEKDKDEKSESEBPKESSGPAEAPAR--TAESELETN-PPEPTGPAAPSPRA	1728	
Db	1645	GDEEK-KVEDSVIEQKEEBAASAPESPRLTPKAPKEENVYKRP-----	1690	
QY	1729	PPTPDTPP-----LRQADER-FDSTLIQTTIPGVALALGSIATFLKKTAKSYGN	1791	
Db	1691	-----PPPKRRIRIKTRNVLDHPAVIPALMSJTIMSIGIPAAIFYFLKKTAKTSYGN	1744	
QY	1782	LFQIQLPKSYDIPRLTKSSNRKIIPYVSDRYKKTITYHMGSD-EDKYAPRSOTDVT	1840	
Db	1745	LFQIQLPKSYDIPRLTKSSNRKIIPYVSDRYKKTITYHMGSSODEKYAPRSOTDVT	1804	
QY	1841	SESEYEELDINDIYVPSPPYKTLIEVLEPPSGNNTJASGKNTPDTRDNDQND---GIP	1897	
Db	1805	SESEYEELDINDIYVPSPPYKTLIEVLEPS-----KDTQNDHINDIPSIDP	1853	
QY	1898	SS---KITDENMOLKEEISMLQNOPDYVNDYTSGNSSTNTNTTTSRHYNDNTN	1953	
Db	1854	NSDTPPTTDEDMQOLKKEIDISMLQNTQTEPN-----ILHDVNDNTH	1898	
QY	1954	TTMSRDNMEENULLPSTHDCNLYSGEYSNV-----MNVN-----SKNDIP	1996	
Db	1899	PTMSHNNDDQKPFMTISLHDNLFSGEYNYDMFNSGNNINISDSTNSMDSLTJNSHSPY	1958	
QY	1997	NRDNNVYSGIDLINDLSGKRPIDYDEVYAKRENFLETENTFKRTSTON--VAKTTS	2054	
Db	1959	NKNDKLVISGIDLINDALSGNH-IDYIDEMLKRENFLETQHHPRNITISNRVYQTSDD	2017	
QY	2055	PIHNDLEFHKWLDHRHDMOKEKKRKNEDJLNLKEBMNKENINNSGKTYNSDKRSHNHV	2114	

Db 2018 PITNOLNLFHKWLDNRHRCMEKKNHNERLPKLELM--EKETHSGDI--NSGIPSGNHV 2073
QY 2115 LNTDVSIOIDMDNPKTKNEITNMDTNOCKSTMDTILDLDAEYNDPNPYVDFEYEDDIITYHDV 2174
Db 2074 LNTDVSIOIDMDNPKTKNEITNMDTNPDKSTMDTILDLDAEYKNEYXYYDFYKHD-ITYDV 2132
QY 2175 DVEKSSMDIIVYDHNNTSNMNDVPTKMHIEMNIVNKKKEIFEEYEPIS 2223
Db 2133 NDDKASEGHIMNDHKMKDNNNSDVPITVQIEANVIN--QELLQNEYRPS 2180

RESULT 5
T286426
variant-specific surface protein 2 - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: T286426
R:Su, X.Z.; Heestwole, V.M.; Wertheimer, S.P.; Guinet, F.; Herfeldt, J.A.; Peterson, D.S.
Cell 82, 89-100, 1995
A:Title: The large diverse gene family var encodes proteins involved in cytoadherence an
A:Reference number: 220487; MUID:95330813
A:Accession: T286426
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2664 <SU>
A:Cross-references: EMBL:LA0609; NID:9886376; PID:9886378; PIDN:AAA75398.1
C:Genetics:
A:Introns: 2197/3
A:Note: var-2

[illegible]

Dd	494	E-LIGNEMKERNK-GECKGGRYNIPIKGRHNHVIPLVSFGDHNKEITE:IEQFC-----	545
OY	585	DGVSIVSVTTGASGINSKEKKELEYDEMKCYKHNHEVOKVVNQGEVEEDDI:TLKAGAGLILP	644
Dd	546	-----AESNDSKSLTEOMKCY-----YGRKEVE-----VCTLE	574
OY	645	NPKKNKEVSEAKSONNHADIQKTFHDFPYWAHMLKDSIMHRTKRLK:GISDCKTMKCR	704
Dd	575	-----EIOKTFHNFPYFMJIRHLLNDSIEWRCK-INICIEKAKEGCK	625
OY	705	NGCKKKDDCEBKVKQKQETMKPIKHNFTQGS--IPBSYFTTIELI:KIQFLKEOTEE	762
Dd	626	NECKTDCGCEQRMWIGKKEEMGEIKHNFPTODGFSIFGNNDPVEYLEVNI:MLDELFODITE	685
OY	763	NTEKS-----LDAEABELKILQ-----KIL	783
Dd	686	AYGSOXQIOGIKPTLAKKTKQAADATBEQKNITIDLLFEYDSEBAKCK:IOBEOQPKPT	745
OY	784	KLEN-----	787
Dd	746	KVRNPDYCGNNTYDALAGKVAQKIQQEAKEQDLDRNDSRSALKMAQSK:SNQGDPPDFK	805
OY	788	-----	787
Dd	806	NLCGITQKHSNAIGDSKNPCNNKKERFYNGEKMKNNGEVMKSHTDLY:JPRROHFCTSN	865
OY	788	-----	787
Dd	866	LEHJNTSTGLTSDKAIHSLGDVLLAAKKEGBDIKTKLTENDNRSSJ:RTMKYSPADIG	925
OY	788	-----ENNL-----	791
Dd	926	DIINGTLDMDIINGDANGVQNNLKDIFS:ITBELKQHPDKFNDNKYV:NDSKHTKLRSDW	985
OY	792	-----	791
Dd	986	WEANRDQVMKAMCTPKNGNIQCGATPHDDYIPORLRMVAEWAEWFC:EOSRLYEELLRD	1045
OY	792	-----AVYNAGTEQ	800
Dd	1046	CGSCTTGKCNNDKCAKCDKOCQOYEYTKTQPMADQNNELSNKQI:LYMOAKIAAINGTEK	1105
OY	801	KTL-----MDK-----	806
Dd	1106	STTTKDKDKNNVIDFQKLHEANYGTRGPPREAHNRDRARRAANSKSDVYETTAGYIHQE	1165
OY	807	-----LINHELNDATYCKD--CPLEBEDKSKGRSADPSPD	840
Dd	1166	ARTRECLGQWVFCNNNNNNEAFSLTPHEYKHAACKENKAKASSPEE--LGRS-DSFDDH	1221
OY	841	FIRPEREKEDENDEDEDVVRD--EENAKETEGSS-----ATDTTSL--DV	885
Dd	1222	QTPREDEVEHSSBEGBEEDSEBEKEEVEEYHOGADEKAGAVSOPEASPTTKDQVVKP	1281
OY	886	CPIMGKVLTKDNESLQDACLKYG--GNNSRLGRCV-----TP	922
Dd	1282	CDIY-KELFSNVDTLQACSTKTKGPPKN--YGRRCIPKITSNDYVTELGOSSRRVYVASTP	1338
OY	923	SGEPTTSDKNGAICVPPRRRLRYLKKIDVMAWK-----TESPCASGSEA-----	967
Dd	1339	-EESGNSDNKNGATCIPRRRRRLRYLVKLEQMANKNHNTETSVSOGCAVTAARSEAPAPAGK	1396
OY	968	-SNSGTSSTPPDSKELMLKAFVESAAIEFFMLNHYKKE-----KAVYQDEAGAGLPRVE	1022
Dd	1397	BSSSGKETPSDK--LRTAFIESAAVEFFPLMDRKKEWLAQKKAELQNE-GADLYSSG	1452
OY	1023	EGSEYEPDEKLKLGKIPDGLFQMFYUULGADYRDILFSGSNDTTSVSH:DTFPSSSNDMLKN	1082
Dd	1453	DGDDP-DPQKKLLNGVYPRPFLRMFTYTGQYRDILLYHGN--TSDGNTNGSNNN---N	1506
OY	1083	IVLLASGSTEQDERKNKKYKEI--KNFRCKSTERSAPMLVSHPQTMWINNGKTYIHWGMC	1140

```
D 1507 IVLEASNGKEDMOKIOEKIEIILPRNGSTPLVPKSS---AOTPDKWNHEASTWKMIC 1563
Q 1141 ALT-----SKDIAGVKEKPOKIEPNELAMDE-----ANKR-----PKPOYOTNV 1183
D 1564 ALTYTEKRPDISARGDE---NKIEKDDEVYKEFGSTADKIGTASTPTGYTKTQYDYEKV 1620
Q 1184 KLDENSGTSPRTTOTQASSDNTPTTLHFVKRPYFEMFEEMGESCEKRRKRLKOIVD 1243
D 1621 KLEDTSG-----AKTPSASDTP--LLSDFLVRPPFYRLLEMGQNFCEKRRKRLKOIEE 1674
Q 1244 CKVENGDVGRCSGGEACDSTHDYSVVPFNCPC---CGHNCSSYRWIERKKEFPHKO 1301
D 1675 CM--DGSCKKTSYSGDEQCD---RRDTSNEVSADLEGRCGNSCREFYKWKIRKREYDQ 1729
Q 1302 SNAYGOQKT---DATRNNG--WTFDEKFTLETWPDAKFLERLKNGCKTNKEYGDDI 1357
D 1730 ANAYSKQTKYEEGSKGGLDHNKEFVCYKIGTCTDAAPLRLKNGCKCKDNENGGNDI 1789
Q 1358 DFEKDSKTFQHTYECGCPKPKFTKQNGNC--GVSGLNGCDDGKSIDAKELAKRSSTT 1415
D 1790 NFGMTEEFRAENCKPCSSFKINCRCNGSCGDDTKEKCNCG--TITGNPNTMGCTE 1848
Q 1416 DVWVRVSNPDNTFEGDGLKACQHANIFKIRKDYMKCGYVCYDICEQTNINERTDG 1475
D 1849 DVVWVSRKNAHEEGDGLDEACENAGIFGTIRKDEMKRCYCGALHCKOKEKGNALDQ 1908
Q 1476 EYIGIRALFKRMVENFLEDYKINDKISHCIKKGSGKSCINGCEKNSCKLEKWIETKIAE 1535
D 1909 QIILVRLALKRWEYFLEDYKIKKIKKLPCLENGNGSTCINGCKKCRVSEWIKLAKDE 1968
Q 1536 WENIKKRENDQYENKDPDYNVKS---ILEELIPKIAVNDQVYIKLCVENSKGCTL 1591
D 1969 WTKIK---NHYLEKKKEGDKWVTSIYTNVLETLVQIAMAANDKRPQTSLDLKTSLGNC 2025
Q 1592 ISNQN--KENDALDCKLKYKAKNCPG-----KPSSEKSDCKEPPRLDEED-Q 1642
D 2026 PENSRKNGNDNDALDCLNLETKRIHECKQIHENSYSNSQPHNCCGNP--DDEEDL 2084
Q 1643 NPEENTLEPPKFPCTPTQPREKGETCGNKEEKKDEKSESEPAKESGPAAEBAPT 1702
D 2085 LEENPVPQGFCTPQOEPP--DCKGKLEKKDEKKEDEPOEPAEDGAI-----VPS 2138
Q 1703 AESEETETNPEPPGTPGAAP--PSTAP--PTPTPLPRLQADEPFDSTLQTTIPPGV 1759
D 2139 G-----PGSEPEADKGPVKAPEIKPOEPDLSHAPV---SLVTSILAMSV 2184
Q 1760 ALAAGSTAFLELKKRTASVGNLFQIIDIPKSDVIDIPLKSNKRIPIVSDRYKATYIY 1819
D 2185 GIGFAAFYFLKRTKSSVGNLFQIIDIPKSDVIDIPLKSNKRIPIVSDRYKATYIY 2244
Q 1820 MEGDSDEDKYAFMSDVTDTVTSSESEYEELDINDIYVPGSPRYKTLIEVLEP----- 1871
D 2245 LEGSGSDS--GYTDHYSITTSSESEYEEMDINDIYVPSPKYKTLIEVLEP----- 2303
Q 1872 -----SGNNTTASGKNTPSDTRNDIONDGIPISSKITDNENQOLKEEISMLQNPNDV 1925
D 2304 GNNNTTASGNNNTTASGKNTPSDTRNDIONDGIPISSKITDNENQOLKEEISQYLQSEPTYE 2363
Q 1926 PNDTSGSSSTNTITTSRRNVNNTNTTMSRDNMENLLPSTHGNLXSGEYSINV 1965
D 2364 PN-----MLGYNVNDNHPPTSHHVEKEPFIHSDNLSSGEYNYDM 2408
Q 1986 -----NMVN-----SMDPIPNRDNNYSGIDLINDSLSGKPIDIYDEVLYK 2028
D 2409 FNSGNPNPINSIDSTNSMDSLTNNHSPYNDKNDXLSGIDLINDALSQNH--IDIYDEMLAKR 2467
Q 2029 KENLFCTE--NTKRTSTQNVAKTTNSDPIHNOLELPHKMLDRHDMCEKMKNKEDILNKL 2087
D 2468 KENLFCTKHTKHTNTYVAKPARDPITQIIMLFHKMLDRHDMCEKMKNHBRPLK 2527
Q 2088 KEENKKNINNSGKTYNSDNKPSHNVLTNDVSTQIDMDNKTKTNEITNMTNDOKSTMD 2147
D 2528 KEIM--ENETHSGDI--NSGIPSGNHVLTNDVSTQIDMDNKTKTNEITNMTNDOKSTMD 2583
```

```
Q 2148 TILDLEKYNDRPYDYFEDDIYHDVDEKSSMDIYVDHNNYTNMMDVPTKMHTEM 2207
D 2584 TILDLEKYNDRPYDYFEDDIYHDVDEKSSMDIYVDHNNYTNMMDVPTKMHTEM 2643
Q 2208 IVNKKKEFEEEPISDIWNT 2228
D 2644 IVNKKKEFEEEPISDIWNT 2664

RESULT 6
T28161
hypothetical protein FC93-va111-1 - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: T28161
R:Hernandez-Rivas, R.; Matted, D.; Sterkers, Y.; Peterson, D.S.; Wellem, T.E.; Scher
Mol. Cell. Biol. 17, 604-611, 1997
A:Title: Expressed var genes are found in Plasmodium falciparum subtelomeric regions.
A:Reference number: 220483; PMID:97154495
A:Accession: T28161
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2647 <HER>
A:Cross-references: EMBL:U67959; MID:91794255; PID:91809295; PIDN:AA047438.1
A:Experimental source: strain FCQ27/PNG
C:Genetics:
A:Introns: 2158/3
A:Note: FC93-va111-1

Query Match 33.8%; Score 4084; DB 2; Length 2647;
Best Local similarity 35.6%; Pred. No. 1.3e-177;
Matches 1036; Conservative 288; Mismatches 632; Indels 954; Gaps 106;

Q 6 GSGGTQDEDAHNVLEFQKVHDEYGEAKNVSELKGSLSLSELTAF--TVKSMQT 63
D 5 GRGT-NKSAKEVLEDEIGETIQKRAHSDADTFRSQIKGN-----FGEAKFYNGEIMQP 57
Q 64 ESKYTEL---LEAN---SKRNPCKKDGNDVDFPSYKEQAGYNNKKMKCSNGM----- 111
D 58 NSKLCEDHTIDTIVTVDGHSNFC---EGQTVFRP-----DDNNSQCTKNRIKOSVD 106
Q 112 ----TCAPFRRLHLCKNFKPMNNSN--DSSKAKIDLAEVMAKYDGEISIKTHPYKDS 165
D 107 NSVGCAPRYRLHLCSNLSIESIOTNNYDSSKAKINLAEVYAKAFEGESIVKNYEDLGH 166
Q 166 KYPSGDEPMCTMLARSFADIGDITIRGDLVYGNKKKKONGKETERELEOKLEIFKKIH 225
D 167 H---TTEGICTALARSFADIGDITIRGDLVYGNPQ-----ESARRKQLEDMLRKIFEKIY 218
Q 226 DNL-----KKKEOKRNGDEDPNFYKLRDMMWTANRETVYWGAMTSKELDSSYPRAT 279
D 219 KELTSSRNGKTNGAIEERYK--DGSQNYTKLRDMMWANNANRLDTWKAMIC-KAPGNAPYFRNT 276
Q 280 CNDTGGQSPQTHNRCRCDKDKGANAGRPKAGDGVTVPTVFYDVPYLYRWFEEMAEDFC 339
D 277 CSN---GEKPTGEKQCI-----DG---TYPTFLDVLVPPQLRLRFEEMAEFEC 317
Q 340 RKKKKKLENEKQCRGKDKSDERYCSBNGYDCQOTISRGKGVPMKGCTDCFPAGCSYE 399
D 318 RKRMLKLNQALNCRGMDDDEKEKYCSRNGYDCTKTRISIDKYSMNEECYKCLVCDPYV 377
Q 400 NWIDNOKRQPKOKK-----YTKELIDGGRRKKRAVGGTTKYEG--YEKSPYEKLKNDGY 452
D 378 KWIDNKKKEFKOKKCCENEIYRNNESSQSPK-----NTNNMETDPTGYLKLKD-Y 428
Q 453 GTVDAFLGLANNEKACKDITDG--GKINFEKYNSSGGVYGGSGGTSAGSGTNDENKSTFY 511
D 429 QSMNDPLKLNLSERTPCNIIDAKSIDF-----TKDPEE--TFS 465
Q 512 RSEYCPQPCDGCYVQHKGNQWER-----KTKVAKMRMSKLYKPING--KMYLLKSLK 562
D 512 RSEYCPQPCDGCYVQHKGNQWER-----KTKVAKMRMSKLYKPING--KMYLLKSLK 562
```

Db 466 HTEYDCPFCWGLKTOADGTMRLYLENDPQCPKPK-----YEPKGVPEPTEDVLYTGK 520
 QY 563 VKDMIIILKKNNKEFCILQNSDSGVSVYTTGASGSEKKELEIDEMKCYHNEVQKN 622
 Db 521 ENKDIIV---KLRECKT---DGNIGF-----KNE---EMWXY---YQVG 553
 QY 623 VQGEVEEDDELKGGGLCILPNPKNKEVEAKSONNHADIQKTFHDEFFYVVAHMLMD 682
 Db 554 NDKCVLENEEELGG-----EKV-----KQDNFLMVAHMLKD 588
 QY 663 SIHMRTKRLKSC1-SDGKTMCRCNGCNKCCDEKFMVKOKETEMPRKIDHFTQESIPB- 740
 Db 589 SIEMRSK-LSNCLKSDKKT--CITTCNDNCQCYDKWIGKAKVHMPIQIKHFPKQTFQGM 645
 QY 741 GYFTTLEILKLOFLKEPTEENTENSLEDAEAELEKLOKLEKLENNLAVNAGTGO 800
 Db 646 GHYF-VLETVL-----BEDQFDTITRAYGDAREIVHJOEMLOKKEEVLH--EDASN 666
 QY 801 KTLMDKLNHELNDATKC-----KDCPL-----PEED--KSNGRSADPS----- 837
 Db 697 KTIIDELDLHELKEAKOCIVNHNKDNCRPADLSDEDEEDIFQRONKCAKPSGTHIRALV 756
 QY 838 ----- 837
 Db 757 NKVASMHKKKROLVNGVSSKLGAADAKGEYRKSGTTIKLKJCSITDHSNAKRGHT 816
 QY 838 -----PDIFP----- 843
 Db 817 DOPCKRKDSKVNVKNRMMWDAGFISNTYKIDYMPRRHOFTSNLEYLOTNNKLNGLND 876
 QY 844 ----- 843
 Db 877 INGNRIINDSLGIVLFRANTEADFIKMYKQNDYKQNDATTCRAMKISFADLDIQR 936
 QY 844 ----- 843
 Db 937 QHICRIMIVERVKEHSEIRNFIILSKNILLAKEIKEDTPYTKLREDWMEANRKKIMPA 996
 QY 844 ----- 843
 Db 997 MOCPTPNPSFPCSYHIGLDYIPORLRMTMAEMFCKEOKROYGELYSASNGKDERV 1056
 QY 844 ----- 843
 Db 1057 KYVRLRVANVOACKHVKIILKMLHIGKEQMDKMEIKYLLILOQTTAANGPRTYSL 1116
 QY 844 ----- 843
 Db 1117 VDENEKPVVNFLELYKENGKIGNPRDTPRAKRSKRETAAPASVAKNDVYSTAAGVHGE 1176
 QY 844 -----RPEEKEDENEDDEDEV 861
 Db 1177 MGRHNECKTQTECEKTEBOYENENTFKNPPOYKADICICNTRPPEKESRKRSEDSDE- 1235
 QY 862 RODETAKET-TEGSAF-----DRTTSLDVCPIYKYLTKNESIDCASCJK 907
 Db 1236 ---EEKVETKEVEKATEDAVDTGPPPAKKEATTLIDVCPYIAGVLTJKN--LENCPPK 1290
 QY 908 YCGNNSRLMGCV-----TPSGEPTTSSDKNGAICVPP 940
 Db 1291 Y-GRAPRTPMKCIPEKTKMAATGSESSSGNGLQRAKRAKTVESGSPVTSN--SGSLCIPR 1347
 QY 941 RRRRLYIKKIIVMAT-----KTESPQASGSEBASSTGTPPDSKEALLAKAPESAAI 993
 Db 1348 RRRRLYIQKILHMAWAGNTVVSQAOTPOGG---TSSPSEKTEPSPDK---LRTAFIQSAAI 1401
 QY 994 ETEFLMHRKKEKKAANOAE-----GAGHGILP--HYEBSPEPYDEDEKIXEKEK 1038
 Db 1402 ETEFLMHRKKEKKAANOAE-----GAGHGILP--HYEBSPEPYDEDEKIXEKEK 1038
 QY 1039 IPDGLRQMEYTLGDIYRDILFSGSNDJTSVSKDTPSSSNDNLKNITLLASGSTEOERERK 1098
 Db 1457 IPPEFLRQMEYTLGDIYRDILFSGSNDJTSVSKDTPSSSNDNLKNITLLASGSTEOERERK 1098
 1457 IPPEFLRQMEYTLGDIYRDILFSGSNDJTSVSKDTPSSSNDNLKNITLLASGSTEOERERK 1497

QY 1099 NKYKEIKNPKRCKSTERSAPNLVSHP-----OTMMENNGKIYMHGWCALTSKDIKANGV 1152
 Db 1498 ---KKIKD---AIEKVLKNAADSQRPSEDEKROTWMQNEHILMNGMIGALVTKREDEKGT 1550
 QY 1153 EKKPOKIEBNIMDEANKPKRPPOYQYNNVKLDENSGSPPTTQOASIDMT---PTTL 1209
 Db 1551 PLK-QNEGKLSALMDEKNNKPKPDQKYQYKVLDENSGSPRIIVPAP:PTTTPPPPPSP 1609
 QY 1210 THFVAKPIYFMFEMWESCEKERRKRLQIYVCKVENDGVKCSGD:EACDSISTHDY 1269
 Db 1610 TSGF-SRPYFRYLEMAEFCEKERRKRLKLEKIVECEWDEDEGKOKCSGD:EDEEIRKODY 1668
 QY 1270 STVPSPNCGCGKHCSSYKMLERKKEIEFHNSMAYGOQKTATATNGN:THFPEKCKTLE 1329
 Db 1669 STVRDQYCEGCKYCRFYKRWIGKKDEYDOKOEAENNOKTARRNN--DNASTYID 1725
 QY 1330 TWPDAAKFLERLKNKNGCKTKNEYGGDDID-----PEKDSKTQHTZYCPCKPKFTN 1381
 Db 1726 TCTTAGDFLOTLKNGPK-----NDNVDSGENKKIFENDDTFRYTJYGCCTSLNGFK 1779
 QY 1382 CQNGMGVSGLNGCDGK---SIDAKETAKRSSTTVYKRVSDNTYFEGDCLKDAC 1438
 Db 1780 CNGDCQVRV-TWVTCNGSNRTTITTAADI-KNGNSABEINMLVSD-DINSGNGFNDLEAC 1836
 QY 1439 QHANTFKGIRKDVWKCQYVCGVCEQOTINERTDQKEIOTRALFKRVENLEDYNTKI 1498
 Db 1837 KNANIFKGIKEKMKCVYFCKSDVGLKKNPD-IDONQIILRALFKFLEFLLDYNTKI 1895
 QY 1499 NDKISHCIGKGEKSGCINCEKNSKLEKMIIEKLAEMENIKKRFNDX:YENKQDPDYNK 1558
 Db 1896 RKKLNPCINNGEKAICTNG-----CYEQMINHKREWMNLKS-FNECY-NGDDTERNR 1947
 QY 1559 -SILEELPKIAVNDQNVKILCVFENSKGCTLSNTQNNK---ENF:AIDMLKLGK 1614
 Db 1948 LRFVLDLIRQIATIDKGNHNGLVKLVKSVKNCNGNSONGKEGBENTLYVLCLOKLEKK 2007
 QY 1615 AKNCPGKR---SGEKOSDCKEPPRLDEEDQ--NPEBNTLEPKFCF----- 1656
 Db 2008 AERKADNPETSGIPQOPCEVSPNHLIEDEQPLEEBENYENHRICTOD LKHNHNOBQR 2067
 QY 1657 ---PTTOPPEKGETCCKNKEEKDEKKESEEPRAKESGPAABEP:PTAASEETENF 1712
 Db 2068 LYKNPLVQPTLKRKKK--KKRKKIKKKKNODFNHRCGAFINMP-----KTKT-- 2118
 QY 1713 PEPGTPAAPSTPAAPTPDTPRPLRQADEPFDTILQTTIPRGVILALGSLAFLETK 1772
 Db 2119 --PPSSG-----KNPMEHPAIVP-----ALVTSILAMSV:IGFAAFYFYFLK 2158
 QY 1773 KTKKASVGNLFQILOPKSDYDIPPLKSSNRYIPYVSDYKKGTYIYIEGSDDEKXAPM 1832
 Db 2159 KTKKSTIDLLSL--IPKSDYDIPKLSPNRYIPYVSGYRKRRIYI:EGDSGTOS-GYT 2215
 QY 1833 SDTTPVT--SSESEYELELDINDIYVPGSPRYKTLIEVLP-----SGNTYTAGSKN-TPSD 1886
 Db 2216 DRYSDTSSSEBEYEMDINDIYVPGSPRYKTLIEVLP-----SGNTYTAGSKN-TPSD 1886
 QY 1887 TRNDIONDIPSSKIYDNEMLQKKEFISNMLONOPNDVNDYTSN:STNTNITTTSH 1946
 Db 2276 TQNDIPTSDTP-PTIYDDEMTLKHDFISNMLONQKQDVNDYTSN:STNTNITTTSHD 2334
 QY 1947 NYDNTNTYKSDNMEENLLPSIHDSNLYSGEYSYNNMYNNSMD:PIINDNMYSGI 2006
 Db 2335 NYDNTNTYKSDNMEENLLPSIHDSNLYSGEYSYNNMYNNSMD:PIINDNMYSGI 2006
 QY 2007 DLINDLSGKPIDIYDEVYKRENELFCTENTKRTSTONVAKTYS:CPJHNOLELFHKM 2066
 Db 2395 DLINDLSGKPIDIYDEVYKRENELFCTENTKRTSTONVAKTYS:CPJHNOLELFHKM 2066
 QY 2067 LDRHRDMCKRNKEDILNKLKEENKKE---NINN--SGRTYN--SDNK----- 2108
 Db 2455 LDRHRDMCKRNKEDILNKLKEENKKE---NINN--SGRTYN--SDNK----- 2108

```

QY      2109 -----PSNIN-----HYLNTDVSLTQIDMDPKTKNET-----TNMCTN 2140
Db      2515 IPSGAGSDIPSDNNHTSHDPIYLNLTDVSTQLHMDPKPLNEFTTYDSNPNOVDPTIYDSN 2574
QY      2141 QDKSTMOTILDLEKYNDPYDYDFEDEDI--LYHDVDEKSSMDDIYYDHNVATSNNDIV 2198
          |||.....|||...|||..|::|||..||::|||..
Db      2575 PONSMTOTILDDLEKYNEERY-----DYQDIYNVN-----DD-NDISTVIIMADV 2620
QY      2199 PKKHIEKNIVNNKKKEIFEEEPISDIWNI 2228
          |:::||:| ..|||.....|
Db      2621 PSKVQIEDMI---NTEIFEETYPISDIWNI 2647

RESULT    7
TI8399
variant-specific surface protein 1 homolog 3D7var1 - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: Tl8399
R:Rubio, J.P.; Thompson, J.K.; Cowman, A.F.
EMBO J. 15, 4069-4077, 1996
A>Title: The var genes of plasmodium falciparum are located in the subtelomeric region c
A:Reference number: Z18927; MUID:96324414
A:Accession: Tl8399
A>Status: Preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-2042 <RUB>
C:Cross-references: EMBL:U53324; NID:g1297090; PID:g1297091; PIDN:AAB09769.1
C:Genetics:
A:Map position: 12
A>Note: 3D7var1
```

```

Query Match 33.7%: Score 4073; DB 2: Length 2042;
Best Local Similarity 41.4%, Pred. No. 2.8e-177;
Matches 951; Conservative 293; Mismatches 581; Indels 470; Gaps 83;

QY 139 HDLLAEVCMARKYEGESIKTHPKYDSKYPGSDPFMTMLRSFADIGDITIRGRDLYIGN 198
      ||| :||:||||:| :| :| :| :||| |||||:||||| ||
Db 12 HNLIDVCLAOYEQSISQNHGHQLSY PDSPSOLCELRASFADIDIDIVGRDLYIGN 71
      ||| :||:||||:| :| :| :| :||| |||||:||||| ||
QY 199 KKKKONGKETEREKLEQKLKEIFKKIHNL--KDKKQKRYNGEDENEFKLRDWMJTA 255
      | :||| :||| :||| :||| :| :| :| :||| :||| :|||
Db 72 NR-----ENDKLEKKLKGFKKIYDNLVEKKKEEAEFTDK -DDADDFQLRDMWAL 122
      ||| :||:||||:| :| :| :| :||| |||||:||||| ||
QY 256 NRETVWAMGTCSEKLSDNSYFRATCNDTGQSPSQTNNCRCDKDKGANAGKPKAGDGVT 315
      ||| :||:||||:| :| :| :| :||| |||||:||||| ||
Db 123 NRQDVMKAITCDA--HBSRYRKMGA-DGSITESAMRQC-----NVA 161
      ||| :||:||||:| :| :| :| :||| |||||:||||| ||
QY 316 IVPYFDVVPQYLRFEEWMAEDFCRKKKKKLENEKCRGDKDSDEYRYCSBNGLDQOT 375
      ||| :||:||||:| :| :| :| :||| |||||:||||| ||
Db 162 DVPNFDVVPQYLRFEEWMAEDFCRKKRKKLKDALQCRGDDGTCKDRYCDLNRDCKRT 221
      ||| :||:||||:| :| :| :| :||| |||||:||||| ||
QY 376 ISRKQVAMGKCTDCEAFAGCSYEMWIDNORRQPDKQ-KYTKETS-----DGG 424
      ||| :||:||||:| :| :| :| :||| |||||:||||| ||
Db 222 ISAKHELVQEGECKCSVYCIPRGFWIDNOKOEFKKQNKNTNEINKKHDETTKEISNR 281
      ||| :||:||||:| :| :| :| :||| |||||:||||| ||
QY 425 RKKRAVGGTTR-YEGEYKSEYEKLEKNDGVTVDALGLNNKCAK---ITDGGKINF 479
      ||| :||:||||:| :| :| :| :||| |||||:||||| ||
Db 282 RKKRSL--TTKNYKYDDEFY-KIKKDEYPRYDKLIDLISKETAGESPURDEPRISINF 338
      ||| :||:||||:| :| :| :| :||| |||||:||||| ||
QY 480 KEVNSGGVVGSGSGTSGASGTNDENKGTFRSEYCORPCDQGVQHNKGQMERKTKVK 539
      | :||| :||:||||:| :| :| :| :||| |||||:||||| ||
Db 339 K-----NYKNPDIFSHPEYCOACPMWCG----- 361
      ||| :||:||||:| :| :| :| :||| |||||:||||| ||
QY 540 KMRMSKLYKPLNGKAVLLKLSLKVVKKDMILLKKNKKEFCCLTONSSDGVSVYTTGA-- 596
      ||| :||:||||:| :| :| :| :||| |||||:||||| ||
Db 362 -----TWYLLMVNVQKQRYAVATIKLFKKWTLIQTNTDIPILPRTDSNIV 408
      ||| :||:||||:| :| :| :| :||| |||||:||||| ||
QY 597 -----SGNSSEKKELVDENKCYKHNVEQKVVNVQGEVEEDDELKAGGGLCI---LPNP 646
      ||| :||:||||:| :| :| :| :||| |||||:||||| ||
Db 409 EKYRNFCHSSSDNNNSDOINNMOCH-----YDESKKSGQNDRCVEGTQWF 453
      ||| :||:||||:| :| :| :| :||| |||||:||||| ||

```

Qy	647	KKKVEYSAKSONNHADIOCTFHDEFFUYVAHMLKOSIMHRTKRLKSCISDOCT-----	700
Db	454	KKDOCT-----SYNAFFKVMSEMLDJSIMKRAELDLCKLNDKTKGKKNY	500
Qy	701	MKC-----RNGCNKRCDFCEKVMKOKETEMKPIKDHFKEJOGEPYFFTELLIK--LQ	754
Db	501	KNCLNFKLNGKLNK-----KKNNGQ-KKHKKKK-----IDKKLVCLLKMHK	543
Qy	755	FLKEDTEENTENSIDAEAE-ELKHLOKILKENENNLVAVNAGTEOKTLMKLLNHELN	813
Db	544	VLANDVFLQ-----DMEKAQODPHIAKIKELLKKNDEKVNNL-SMERTIFDOLLQEBQ	597
Qy	814	DA-----TKCDCLPREDKSRGASADPSDITIPREPEKEDDENDEDEVEDRDE	865
Db	598	DAOKVSNNPKECSEETQKPLTDGAPG-GAPRSPDGTIDNLEBIDISDGREDDVSHVEE	656
Qy	866	ETAKETTEGSA-----TDPTTSLDVCPIYKVLFDONSLODACL	906
Db	657	EPEDNPVYEGSEEBEKOEYVAKOTEPANPKDQPEEVNCKTY-ELERKSTNFEADGL	715
Qy	907	KYGGNNSRLGRCV-----TESGEPTT-----SSDKNGLCIVPRRRRLYK	948
Db	716	KYGNK--YGMKCVCHNSDKGSEPTLARGHSHVARSADGAPSGDKOALCIPRRRRKLYH	772
Qy	949	KIVDMATTESPOASGSSEASTSGSTTPDPSKALLKAFVESALIEFFLHMRYEERKA	1008
Db	773	KI-----EGVDTTD--KSLKMFIESANVETFLAMRYK--KLI	808
Qy	1009	VAOEGA-----GHLPVYEGSPREYDEDKLGGKIPDGLRMFTLGDYDILFS	1060
Db	809	TPQSSPLLGWITGVG--VENGDDENNKELLQGEIPDGLKRMFTLGDYDILFS	865
Qy	1061	GSNDTYSKQTPSSSDNLKNIYLLASGTEQERKKMKYE-----TKNPKCSTERS	1115
Db	866	GDKOKKNYSIDYSGD-----VI-----KREMTIEKIASFFQONKKGSTPHV	910
Qy	1116	APNLVSHQPTWENNKGKIMHGCVALTSCDKTAKAVEKKPOKIEPENLME-----A	1168
Db	911	PKNPQTPQTMKNNAKHIMGHICALTYEETKSGSDGK--KIEKDVAUYKKFPGTPNG	968
Qy	1170	NKKPKP-----POXYUTNKL-DEMSG--TSPTTQTOQASSDNTPTTLNHFKR	1215
Db	969	NPLPQPGTNGTSNBPISOYQIDOVVLKEENNGMSTSPKST-SAABSQDTPPTLLQFVLR	1027
Qy	1216	PTYFRMEGESEFCEKRRKRLKQIKVDCVENGVDYGRSGGGEACDSISTDYSTVSF	1275
Db	1028	PTYFRYLEEMQONCKRKKRRLMOIYEDC--GNKXVCSGGDCEYERKQDYKISNF	1084
Qy	1276	NCPOGGRKHSYRKWIERKKIEPHKOSNAYGOQKT-ATRNNG-----NFDKEFCCTL	1328
Db	1085	NCPOGGRKHSYRKWIERKKIEPHKOSNAYGOQKT-ATRNNG-----NFDKEFCCTL	1144
Qy	1329	ET-WPDAKFLERLKNPC-KTKKEVGGOIDDEKXSKTFQHTYEGVGPCKRTKTCOON	1386
Db	1145	GTDVASIESFELEK--GSCSKNNKNDGDTINKPEBDVTFKPADNCKPCSEKVCACENGK	1202
Qy	1387	CGVSGGLNGCDSIDAKIEIAXMRSTTDVYVARSNDNTNFEGBDKLDAQOAHNIFKG	1446
Db	1203	CSGGKKKYNCKRKTATIAELAMINISTBEVFMTVSDNSDHFPEBG--LEPGSANIFKG	1260
Qy	1447	IRKDVWKGIVYGVDCIQTNINERTDKEYIOIRALFRWVENPLEDYINKINDISICI	1506
Db	1261	IRKDVWKGIVYGVDCIQTNINERTDKEYIOIRALFRWVENPLEDYINKINDISICI	1320
Qy	1507	KKGEGSKINCCEKNSCKLEKATIEKAIEMENIKKRFNOYENKROPOYNVKSIIE----	1562
Db	1321	RKGERJTICIN-----DCEVKMINIKKKMETIREXYVQOYTTGSHDITKYVTSFEDPOF	1374
Qy	1563	--ELIPKIAVNDODNYIKLCEVFNKSGGLTISDTONKEN--DAIDCMLKLKYGAKNC	1618
Db	1375	HNEVLAKAIKPGCDLDK-----FQNSTDCTVYASSNSNGYTNKKDIVECLLEMLKTKATC	1428
Qy	1619	PKRPSGEQSDCKEPPLPDEEDONP--EENVTLEPFCBPPTQPREKGETCGNKE	1674

Qy	116	APNLVSHQPTWEMNGKRYIHHGWCALTSKDIANKVEVKPKIENPEJMADE-----A	1168
Qy	911	PKNPVQPTQPMKMAHIMHGMICALITYEKKSSGDGK--KIEKDAAVYKFKPTGONG	968
Qy	1170	NKKPKP-----BOYOYTVNKL-DENSG--TSPTTQTOASSDNTPTTLHFWK	1215
Db	969	NPLPQPGNNGTSNEPIISOYQYDQVVLKEENNGAMSTPSKST--SAAPSDNPTTLQPVLR	1024
Qy	1216	PTTYRMEWGESCESECRKRRKRLKOIKYDCKAVENGVDYGRSGGSEACDSISTHDYTSVSEF	1275
Db	1028	PTYRYLEWQONCKCEKRRKRLQOYEDCR--GNMKVSGSGGDECEBVRKODYEKISNF	1084
Qy	1276	NCPPGGRKCSYRKYERKKIEFHKOSNAYGOOKTD-ATRNNG-----NTEDEKFCFTL	1328
Db	1085	NCPPGGRCKRYKKWMTTKTKEDEQKKAIVDRKKDKYKTEENGAESKHSISDQNFVKKL	1144
Qy	1329	ET-WPDAKFLERLKNRPC-KTNKEYAGDIDREKDSKTFQHTTEYCGPCPKRYTCQONG	1386
Db	1145	GTDVASIESFLEKL--GSCSKNNKNDGDTINKRPEDVYFPRADNCKPCSEKXKVCENGK	1202
Qy	1387	CGVSGLNGCDDGKSDIAKEIANKRSSTJDVYKRVASDNTNTEGDDLKDACQANIFKG	1446
Db	1203	CSGGGKKVNCNRKNTIATEIANNINSTEGEPTVASDNDSHKFEFG--LEPGSANIFKG	1260
Qy	1447	IRKDYMKCGYVCYDICEQJNINERPDQKEYIOIRALFLRWYENLEJYDNKINDISICI	1506
Db	1261	IRKDYMKCGEYGVADICEPTFPFGKONGKEYIOIRALFLRWYEFLEJYDNKIKHISICHM	1320
Qy	1507	KKGSGSCINCCENKSNCKLEKWIETKIAEMENIKKRFNDOYENKQPODVNVYSILE----	1562
Db	1321	RKGEJTJCIN-----DCEKWMJNKKKMDETIREKYVQOYTTGSHDITKYTSFLEDPQF	1374
Qy	1563	--ELIPKIAVNDQDNIKICLVFENSGCJLISNTQNNKEN--DAIDCMLKLKGKAKNC	1618
Db	1375	HNEVLAKAIRPGCGLDK-----FQNSTDCYVAVSSNGYTNKKDIVECLLEHLKTKKATC	1428
Qy	1619	PKRPSGEQSDCKEPPLPDEEDONP--EENVTLEPFRCEPTTQPPREKGETCGNKE	1674

1429 PNOVGENEFT--CDSLPHVDEDDDEELEEETEEVTEVQPIPCQLKRPFPQPG----- 1479
1675 EKKDEKKEESEEPANEESSPAEAPPTAESEETETNPPEPGTGPAPSPPTPTPT 1734
1480 -----DEDCKQASPAF---SECTEN---QPPVIRKEEAPAPRPRPRPR 1518
1735 PPPLRQADBP--FDSTLIQTITPPGVALALGSIAPLAKKTKASVGLFQILQPKSD 1792
1519 KPPQEQPPLPPALKNAMLSSTIMSIGIFATFYFLYKKRKSXSVGLFQILQPKSD 1578
1793 YDPLTKSNRRIPYVSDRKGCTTYTMEGDEDEKFAFMSPTDVT--SSESEYEELDN 1851
1579 PDLPTKLSNRRITPTSGYRAKRIYILEGDSGDS--GTTDHYSDTSSSESEYEELDN 1637
1852 DIYVPSPKRYTLLEVLEP-----SGNNTTASGKNTPSDTRNDIQNDGIPSSKITDN 1904
1638 DIYVPGTPKRYTLLEVLEPSPGNNTTASGNNTTASGKNTPSDTQNDIPSGDTPNNKLTDN 1697
1905 ENNOLKKEETISMLQN--QPNVDPNDYTSGNSSNTNTTTSKRHNDVNTNTMSRDNME 1963
1698 ENNLTAKDEFISMLQSEQKDPVNDYSGDIPFNTQ-----PSTLYEPDNQE 1744
1964 NLLPSIHGNTYSGEYEYNNM--VNSMNDIPINRDNVYSGIDLINDLSGKRPDIY 2022
1745 KPFITSIHRODLTGEYEYNNMSTNSMDDIPISGKNYVSGIDILM--TVCGNHVVDLY 1803
2023 DEVLAKKENELFGE--NFKRTSTQNVAKTTNSDPIHNLQLEHFKWLDHRHDCERKKE 2081
1804 DELLAKKENELFGEKHTHTYTVYNAKPARDDPLHNLQNLFPHTWLDHRHDCERKKN 1863
2082 DLNLTAKKEE-----KKNENINNSGK-----TVNSDKRP----- 2109
1864 ERLAKKEEMENDTSGSKHSDIPSGKLSDPSPDNHISDIPSDIPSGKLSDIPSDKY 1923
2110 ---SHNVLNTDVSIQIDMNPKTKEI-----TNMDTNDKSTMDTILDDLY 2155
1924 CCSSNKTLTWDTYSIQIDMNPKTITNEFTYVDSNPNOVDOTYVDSNPDSMTIILEDLY 1983
2156 YNDPYVDFEEDDIYHDVDEKSSMDIYVDH--NNVTSNNMDVPTKMHIEKNYNNRK 2213
1984 YNEP--YVVDLDD--KYVDVN-----DHDASTVDSNNMDVPSKYOLEMD--VNFR-- 2027
2214 ETEFEETPISDIWNI 2228
2028 -LVKEKYPIADLMDI 2041
RESULT 8
728625
variant-specific surface protein 3 - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: T28625
R:Su, X.Z.; Heatwole, V.M.; Wertheimer, S.P.; Guinet, F.; Herfeldeit, J.A.; Peterson, D.S.
Cell 82, 89-100, 1995
A:Title: The large diverse gene family var encodes proteins involved in cytoadherence an
A:Reference number: 220487; MUID:95330813
A:Accession: T28625
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3006 <SU>
A:Cross-References: EMBL:L40609; NID:g886376; PID:g886377; PIDN:AAA75397.1
A:Genetics:
A:Gene: var-3
A:Introns: 2597/3
Query Match 33.0%; Score 3989; DB 2; Length 3006;
Best Local Similarity 32.0%; Pred. No. 3, 1e-173;
Matches 1034; Conservative 315; Mismatches 644; Indels 1236; Gaps 102;
QY 8 GGTQDDAHNVLDPEGQKYNDEVHGAKNYVSELKGSLSLASTIGTATYVKS---MOT 64

6 GGT--KTKAEVLDDEICKVEQKATEBDALTYRNDLQNSLQAFHG--VPIDVKNPCDLNTE 63
65 SKYTELIEANSRNPCKDKGKNDVDRFS--VKEQAGYDNKKMKCSNG-4-TCAPFRLHL 121
64 I--HTVNVKRRKKENPC-----RGREKRFSVDLSGCAKNTIDSVTNS/GACAPFRLHV 118
122 CKNRFPNANSNDSKAKHLLAEVCAAYEGESIKTHPKYDSKYPGSDPFCMTLAAS 181
119 CDRNL--ELIKTEBNTSTHDLVDVLTAKHEGESLYKHKEYIK--NNFNICYVLARS 175
182 PADIGDITGRDLYGNKKKKONGKEFEREKLEOKLEIFKTHIDL-----KXK 231
176 PADIGDITVKGDLTYLGYDEEKE-----KKOLEKNLKKFFQIHDDVMTSGRTNKKSA 231
232 EAQKRYNGEDENFKYLRBDWMTANREYVWAGTCSKELDNSSYFRTCN-----DTGOG 286
232 EAQKRYN--DATGNYYTLRBDWMNANRDYWKAITCD--AADNDEYFENS SDGLYVFSNG- 288
287 PGTNKKCCGDKDKANACKPKAGGDVTVTPYTPYVQYLRWPEERAEDECRKKKKL 346
289 -----C-----GRNECK-----VPTNLDYVQHLRWPEDEMAEDCCRKNITL 325
347 ENLEKOCRGKDKSDEYRYCSRNGYDCEOTISRKGVRMGKCTDPCFPGSYENMTDNOR 406
326 KSAKEKCGEGNT---KYCSLNGYDCTKYFEKKDKSCSSGNCCTAGSNJCIAIDAMLNOR 382
407 KQFDOK--KYTEISDGGGKKRAVGGTTYEG-----YER-SFEYELKNDGYG 453
383 NEFEKKIKYTFEIE-----KYSSDSKSNISNKKYHIEYENFGKEYE 428
454 TYDAFLGILANKEACDITDGGKINFEYNSGGGVGGSGTSGAS/YNDEKKGFTYS 513
429 TLQNFPLKLNKMYCOE-----KIEEEV-----IIPNKEDVYFHS 466
514 EYQCPPCDGYOHKG-----KIEEEV-----IIPNKEDVYFHS 529
467 EYQCPPCDGYOHKG-----KIEEEV-----IIPNKEDVYFHS 526
530 -----KIEEEV-----IIPNKEDVYFHS 537
527 EYFCSSTYNEGRNVQKWKCYKNKSDNYNCENMISSYKSDTANVYLIVECFHSMKNLL 586
538 VKKMRAS--KLYPINKKMYLLKSLKYVD-----NOWERKTK 566
587 IDTIKWEHQLKNCINNTNTYTCES--KICINCEYERKIRKEHEWEK/ANVEGNNNRMSY 645
567 -----KKNKEF-----KKNKEF----- 577
646 IYYNNLSRVDFSEFLFOYVAFALDODEKGWDOFTEDLKKRFPESKTNTPYKGSODATIEPLL 705
578 -----CLTQNSSDG--SVGSVYTTGASGENSEK-----KELYDEMKCYKHNENYOK 620
706 DHIKQDALCLCRNNNSNESCDVSKKVTNPGKNPSASNNLVKRLAEMQRYARQOLEK 765
621 ---VAVGE-----VAVGE----- 626
766 RGEELINKGDAKGYRROGAPADGFKNVCINQNTHTVONNNRATYVJGPTCTGKDSNGC 825
627 -----VEEDDELGAGGIC----- 642
826 VRMKIGTPMKPGROIOMSAEDIYMPRRONMCTSNLEYLTQYDGPLKGGDKLVNNSFLG 885
643 -----LPNPKKNKEV-----LPNPKKNKEV-----S 653
886 DYLLSAKMDAGKIIELYKKONKNSNLTPEDEDESACRALRYSPADGLIIRGRDLMDKNS 945
654 EAKS--ONN-----HADIOK-----TFHDEFT--YVVA--HMLKSIHRTKR 690
946 DAKRLQTNLKEITFKIKEELPEDIKKKYDKDTHKILREDWMEANFHOV-----WRA-- 998
691 LKSCISDGKTMCRNG----- 706

Dh 999 MCALENDKDMKC-NGIPEDIYIPQRLRMWEMAEWFCEDSRLYNKLVLADCKSGKRAK 1057
Qy 707 -----CNKKDCFEKVVOKETEMKPIK----- 729
Dh 1058 SCTOKDGCOTCKACADYNNKKIKRMEOWEKIKKKYAOQLYKALDYNKGEESKKKTAS 1117
Qy 730 ----- 729
Dh 1118 DAKDOVVHFLAELIRKSGGKGNVKTIVSPPTTPTLYSSAGYIHHELGRTVGCNT 1177
Qy 730 ----- 729
Dh 1178 QKEFCYKNGKAYAKDPKGYEACCKDNRPQPARKEDEDACDYVYLLKDKGETD 1237
Qy 730 ----- 729
Dh 1238 DIDGCONYKAKGDKYPGWDCNSQIHTTHNGACMPPRQKCLCVSGLTJTDRIKAIETRT 1297
Qy 730 -----DHFYTOEG-----IPRGY-----YFT----- 745
Dh 1298 EFIKSAIETHFAMDRYKEDNGEAEALKNGINPEGFKRWYTFGRDIFFGDISIH 1357
Qy 746 -----TELILKLOFLKEDTEENTSL-----DAE 771
Dh 1358 AYISGVSPKVTILEKENDAYAAKQNSNELLDDWMDHQKDIWGMICALTHKISDEE 1417
Qy 772 EAEELKHLQKILKL----- 785
Dh 1418 KKEELKNNYSYKKLINESPKGSKNVDFPAKKPOLFWFIEMGDEFCAQEBEKAAYKVS 1477
Qy 786 -----ENENLAVNA-----GTEOKTLMQ-----KLNHE----- 811
Dh 1478 DAKDYGCCKNTKSNASCYKVEDYITKKKVEYTKQKGFDAKIDTDEGEGFSTKD 1537
Qy 812 -----LNDATKCK-----DCPLBEE-----DKSRGRS 833
Dh 1538 ASEYLLKKKCLDDTCNCQKVKNNTEYMTPKNTYTNSMLEKRCBCQPOEPPEPPREGAR 1597
Qy 834 ADPSHDIFPRP-----EKKEDBN----- 853
Dh 1598 SDSGRDTPPRAGSDASNYPSPRRPADYHEVAEVEGEDEDEDGOLPEQODEV 1657
Qy 854 -----EDDEDEVD----- 863
Dh 1658 EVAGAEEDLDVGAARILGRRTNSPDEDEDEASEEDDDDDADQTEVAGQGEETA 1717
Qy 864 -----DEETAKETEGSA-----TDTTSLDVCPIYKGLVLDNSLOACSLK 907
Dh 1718 DHODTTEETVQERKAEBKDGGETPOKETOPKVEVNPCLIV-KTLFTTTETLKACPTK 1776
Qy 908 Y-GGNNSRLGRCVTPSGEPTTSSDKNGAICVPPRRRLYIKKIVDMATKTESPQASGE 966
Dh 1777 YVNGEKPRPNMKCI-----SSGSDASGSIICPPRRRLYLAKI-----EGVD 1818
Qy 967 AS-STSGSTTPPDSKEALLKAPVESAIETFFLMHRYEKKAAVQEGAGHG--LPRAVE 1022
Dh 1819 TTVSSDGETTTPITHDALREAFIQTAAVETFFLMHRYKIKKEKEQOEELQNGTFLIPPAQ 1878
Qy 1023 EGSPVEUDE---DKLKEGKIDGFLROMFYTLGVDRIFFSGSNDTIVS--SKOTPPSSN 1077
Dh 1879 KVSPEQNEHQKLLKEGKIDGFLROMFYTLGVDRIFFSGSNDTIVS--SKOTPPSSN 1077
Qy 1078 DNLKNIVLASGSTEQERKKNKKYKAIKFRKCSYERSAP-----NLVSHQPTWENNKG 1132
Dh 1939 KNINDI-----SEKISYIEKSEGOJTPPGPKPQGTITTKPEEMWQKNKE 1981
Qy 1133 YIHGMVCAAL-SKDKAKGYEKKPQKTEPENIDAEANKPKPPQOYOTYVVKLDENSGT 1191
Dh 1982 HINAMICALTHNTDROVDQVKGQLEFENGKN-----TPKNSQOYKYNV----- 2026
Qy 1192 SPRTQTOASSDNTP---TTLTHFYKRTYFRWPEMGESGCREKKKLLKQIDKCKEYEN 1248
Dh 2027 ---TIVSSVSSNGGPIGNIKLEQFASRPTPLRWLEWGEFCRCROYHKLRIKECHKDG 2082

Qy 1249 GDVGRCSGDGEACDSISTHNDYVTPSFNCPGCGKHCSSYRKMIERKKIEFHKQSNAYGOQ 1308
Dh 2083 N-RNCDDDGFECKMCMCKCKGDSFEETLCKPSCAKSCAKSYKKKWKISKKDEPFRKQKAYIKQ 2140
Qy 1309 KTDATRNNGNTFDEKFCCTLETPDAKFLERLKNPGCKTNKEYG-GDDIDFEKDSKTFQ 1367
Dh 2141 KKDAGGNNDY--KEFSKTLRNYNDAALFNLKSLKPCSKNDSDVQOEIKFDDERKTFG 2198
Qy 1368 HTEYCGPCPKERTNC-QNGNCGVSGLNCGDKSIDAKELAKRSSTTDVVMKRSNDT 1426
Dh 2199 HETCYKPCSKITVCKCKENHNDNSKN--DCRNINSISADEIKRSNSQDVTMVSYSNT 2257
Qy 1427 NTFEGDDLKACQAHANIFKIRKIDWKGYGCVGIDC--EOTNIEPRTDKE-----YI 1478
Dh 2258 NGNKFTYDLND-CIRKAGITKGIREDYWKCGEYCGVDICTLEKTNNBERVSAKENDKNQIT 2316
Qy 1479 QIRALFKRWENFLEDYKINDKISHCIKGGSKCINGCEKNSCKLEKWKIEKIAEMEN 1538
Dh 2317 LIRVLFRWLBSFLEDYKINDKISHCMKNDKSPCINGCQNKCKVCYKIEKKKSEWGK 2376
Qy 1539 IKKRPNDQYENKQPD-YNKSILEELPKIAYVNDQNVIKLVFENSCKGTLSINTQ 1597
Dh 2377 VREYINQYRDKNSNEAEVKSFETLLPQIPVYTDKGRHDSLTQKLKLCNCSEKSEN 2436
Qy 1598 NKENDAIQCMCLKGLVAKNCPGKPSGKQSDCKEPPLPEEDQNP-----EENTL 1649
Dh 2437 SNEKDVYVCLKLKLEDKAKNKQDQASG-----PCPQTSSENPDDEDILLEEKNPV 2487
Qy 1650 EPPKFCPTTOPPEEKGETCGNKEEKKEKKEESEEPKAEESGPAEAPAESEETE 1709
Dh 2488 EAPNICPKVEEP-----EPVVEEKKDLAEAPSKESS-----TEENSSEG 2527
Qy 1710 TNEPPEGTGAAPSTPAPPTPDPPPLRP-----QADEPFD--STLTOTTPRGVLA 1762
Dh 2528 SNSEQNPBSKPEEPPPTSETPTPPAPPTIOPSQADQPTNSIDLSLTIPGALIA 2587
Qy 1763 LGSIAFLKRRKTKYASVGNLFOILOIPKSDYDIPLKSSNRYIYVSGRYGKTYIYMEG 1822
Dh 2588 LTSIYFLFKKTKTSSV-DLRLVNLIPKGEYGMPLTKSSNRYIYASDRYRGKTYIYMEG 2646
Qy 1823 DSDDEKVAFMSDQTDVTSSESEYEELDINDIYVPSPKYKTLIEVLEPSCGNNNTASGN 1882
Dh 2647 DSDSGHY--YEDTJDVTSSESEYEELDINDIYVPSPKYKTLIEVLEPSCGNNNTASGN 2704
Qy 1883 TSPOTRNDIONDGIIPSSKITDNEMNOLKEFISNMLOK-OPNDVNDYVTSNGSSNTNIT 1941
Dh 2705 TSPOTNDIONDGIIPSSKITDNEMNOLKEFISNMLOK-OPNDVNDYVTSNGSSNTNIT 2764
Qy 1942 TTSRHNVDNNTNTMSRDNMEENLLPISHDGNLYSGEYSYNNVMVSNMNDIPINRDN 2001
Dh 2765 TTSRHNVDNNTNTMSRDNMEENLLPISHDGNLYSGEYSYNNVMVSNMNDIPINRDN 2001
Qy 2002 VYSGIDLINDLSGKRPIDYDEVYLKREKNEFEGTENTKRTSTONVAKTNSDPIHNOLE 2061
Dh 2812 VYSGIDLINDLALNG--DVYDIDEVYLKREKNEFEGTENTKRTSTONVAKTNSDPIHNOLE 2061
Qy 2062 LFHKWLDHRDMCKEKKKEDILNKIKKEENKKNINSNGKTYNSDNKPSHNVINTQVSI 2121
Dh 2870 LFHKWLDHRDMCKEKKKEDILNKIKKEENKKNINSNGKTYNSDNKPSHNVINTQVSI 2121
Qy 2122 QIDMDNPKTKNEITNMQDNQKSTMDTILDLLEKYNPYPYVDFEDDIYHDVDEKSSM 2181
Dh 2924 QIHDNPKRINPEFN-----MDTILDLDDLINNE-YDVQDD--IYDVAN----- 2965
Qy 2182 DOIYVDH--NNVTSNMDVPTKMHIEKNVYNNKKEIPEEYPISDIWINI 2228
Dh 2966 ---DHDTSTVDNSNADVPKSVQIEMD-VNTK--LVKEKYPIDVYDI 3006
RESULT 9
T18378
variant-specific surface protein 1 - malaria parasite (Plasmodium falciparum)


```
Qy 1569 ----- 1568
Db 1711 TYENVNINNCKADONERFKIGOKMFKYIGIRKOLCIPPREHMCDDLSMLGRTT 1770
Qy 1569 -----AVNDQDNVILCVFENS----- 1566
Db 1771 ISDSSALLKRIQEAASERDIIIRKLEONSCDEHRICDAMKYSFADIGDIIRGRDLNWK 1830
Qy 1587 -----KGC-----TLISNTONKEND----- 1602
Db 1831 NSKOKGLOKLEAFINITYNKLONDKNKYEKOPRYLOLRSDMWDANKHIMAMTCNAP 1890
Qy 1603 ----- 1602
Db 1891 DDAFKLKNPNDTSGSSSKGIMTTHSNCGYDKPEPDYIIPQPFMMQEMSFCRLN 1950
Qy 1603 -----AIDCKLAKLYKAKKPC----- 1620
Db 1951 EEMQEPKTCGECKKNSITCEDDRNGTNCENCKNOCEKYLKLIHNMKLGFDKYEIYNEI 2010
Qy 1621 -----KPSGEKOSD-CKE----- 1632
Db 2011 YNNKDSKINSNEYFKKLEKIKDKCKELNSDKCIDETHTCTKTFKSNSKNHNNNTAFK 2070
Qy 1633 -PP-----PLPDEEDONPEE-----NTLEPPKFC----- 1655
Db 2071 NPPEYKACKCDAPDPLDNCPKDSATYERKACNTLPTKLCESKTFNNDDSDMTSFYQT 2130
Qy 1656 -----PP----- 1657
Db 2131 SPRDNTGVLVPPRRROICLNKNTTKLSIEKIDPFKAELMTSAVNEGKLLCELYKKDRDV 2190
Qy 1658 -----TTOPPEKGG----- 1667
Db 2191 TLQAMKSYFYDYGIVGTDLISTAPLDKLTKLNLVKGDTNEIKEDRCKMWTENRTR 2250
Qy 1668 -----ETGCKKEKKD-----BKKESEEPKAEESG-- 1693
Db 2251 VMHAMLGKYKAKGIEBDCSLPDNDHQFLRMFREWSEHFCAKROGLFNEVREKCSA 2310
Qy 1694 -----PAEEFAP-----TAESE-----TENPEPPEGTGAAP----- 1723
Db 2311 QCILEYGTIDPPVCEACTOYRDYITRKIOYRLNLYOYNNEFKAEVTAKEYEYENDK 2370
Qy 1724 -----PSTP-----APPTD 1733
Db 2371 CNOKCNLSKYIDIEKKMKNYDSFDNDLKNKCICQIKRPPKPKYKPEEHTPESQD 2430
Qy 1734 TPEPLRPO-----ADEPEDSTLQTTIPGVALALGSIATFLKKTAKSVGNLFOIL 1786
Db 2431 TPEPLRPPDLPPEAEPEFRDILEKTIPIPGIALALGSIATFLKKTAKSVGNLFOIL 2490
Qy 1787 QIPESDQDIPTLKSNNRYIPVSDRYKGYIYMEGSDDEKAFPMEDTIVT--SSSEY 1845
Db 2491 HIRPSDQDIPTLKSNNRYIPVSDRYKGYIYMEGSDDEKAFPMEDTIVT--SSSEY 2549
Qy 1846 EELDINDIYVPGSKYKTLLEVLLEPGNNNTASGKNTPSRTANDIONDGPPSSKITDNE 1905
Db 2550 EELDINDIYVPGSKYKTLLEVLLEPGNNNTASGKNTPSRTANDIONDGPPSSKITDNE 2609
Qy 1906 WNOJAKEPISMLONOPNDVNDYTSNGSNTNTITTTSRHVDNNTNTTNSRDNEBNL 1965
Db 2610 WNTJKEDEFISMLONONEPTEPN-----MLGYVNDNNTHTPTTSRHNBEKP 2654
Qy 1966 LIPSIHGNLYSGEESYVNVAMVMSMDIPINRDNNYSGIDLINDLSGSKPDIYDEV 2025
Db 2655 FINSIHDRDLYSGEESYVNVAMV--NDIPISARNGNSGIDLINDLSNSK--VDIYDEL 2711
Qy 2026 LKRENELEFGEENTR--TSTONVAKTNSDPIHNOLEFHWLDRHDMCKEY--KNKE 2081
Db 2712 LKRENELEFGEENTR--TSTONVAKTNSDPIHNOLEFHWLDRHDMCKEY--KNKE 2771
Qy 2082 DILNKLKEWKNK-----NINNSCKTYSNDKPS-----HN--HVLNTDIVSIQ 2122
```

```
Db 2772 ELLDKLEWKNKDNNSGINSNTPTSPDIPSKQSDIPEDNNIHSDIPVLTVDYSIQ 2831
Qy 2123 IDMONPKTNEITMADTNOQKSTMDTJIDLEK--YNDPIYDFEDDIYHNDUYEKS 2181
Db 2832 IHMONPKTNEITMADTNOQKSTMDTJIDLEK--YNDPIYDFEDDIYHNDUYEKS 2181
Qy 2182 DDIVVDH--NNVTSNNMDVPKTHIEMNIVNKKKEIFEPEEYPISDIWN 2228
Db 2884 -----DHDSTIVDTNADPEKVOIEMD-VNFK--LVKEKPIPLDMDI 2924
```

RESULT 10

variant-specific surface protein 1 homolog PFB0010w - malaria parasite (Plasmodium fa
C1625
C:Species: Plasmodium falciparum
C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C:Accession: C71625
R:Gardner, M.J.; Tetteijn, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.
; Pereira, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H
Science 282, 1126-1132, 1998
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600; MUID:99021743
A:Accession: C71625
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1711 (GAR>
A:Cross-references: GB:AE001366; GB:AE001362; MID:93845070; PIDN:MAC71792.1; PID:9384
A:Experimental source: clone 307
A:Genetics:
A:Gene: PFB0010w

Query Match 27.3%; Score 3297.5; DB 2; Length 1711;
Best Local Similarity 35.6%; Pred. No. 3.9e-142;
Matches 832; Conservative 265; Mismatches 502; Indels 741; Gaps 87;

```
Qy 1 MATSGSGGTQDEDAKIVLDFGQVHDEVGAEKKNVSELSLSTIGETAFYVKS 60
Db 1 MATSGSGGTQDEDAKIVLDFGQVHDEVGAEKKNVSELSLSTIGETAFYVKS 60
Qy 61 MOTESKYTELLEANSKRNPK-----KDGKND-----VDRFSKEQAGYNNK 104
Db 61 MOTESKYTELLEANSKRNPK-----KDGKND-----VDRFSKEQAGYNNK 104
Qy 49 TYRNDKHP-----EGSTENPNCKLQYDYNTNWTYHGQEPCTDIVERFSPTEGACQCKK 105
Db 49 TYRNDKHP-----EGSTENPNCKLQYDYNTNWTYHGQEPCTDIVERFSPTEGACQCKK 105
Qy 105 MKCSGNGTCAFERLHLCNKNFPMNNSDSKAKHDLAEVCMAXYEGSITKHYPKYD 164
Db 105 MKCSGNGTCAFERLHLCNKNFPMNNSDSKAKHDLAEVCMAXYEGSITKHYPKYD 164
Qy 165 SKYSGDFPMCTMARSFADIGDIIRGRDYLGNKKRKGKQKETEREKLLEQKLEIFK 224
Db 165 SKYSGDFPMCTMARSFADIGDIIRGRDYLGNKKRKGKQKETEREKLLEQKLEIFK 224
Qy 166 ETNPDTKSQCLTVARSPADIGDIIRGKDLRGNTREK-----KRRKLEENLKTIGHI 221
Db 166 ETNPDTKSQCLTVARSPADIGDIIRGKDLRGNTREK-----KRRKLEENLKTIGHI 221
Qy 225 HDNLK-----DKAQRKRYNGDEDPNFYKLRDWTANRETWGCAMTCSKELNSSFRA 278
Db 225 HDNLK-----DKAQRKRYNGDEDPNFYKLRDWTANRETWGCAMTCSKELNSSFRA 278
Qy 222 YDELKNGKTNKEELQKRYRDKDNDYOLREDWMDANRETVMALICN--AGSYQYQSP 279
Db 222 YDELKNGKTNKEELQKRYRDKDNDYOLREDWMDANRETVMALICN--AGSYQYQSP 279
Qy 279 TCNNTGGPQSTHNRKCGDKDKGAKNAPKPRAGDGVITVPTPYVPOYLRMPEWAEDE 338
Db 279 TCNNTGGPQSTHNRKCGDKDKGAKNAPKPRAGDGVITVPTPYVPOYLRMPEWAEDE 338
Qy 280 TCG-RGELPYVTLKCC-----TAG-----VPTPYVPOYLRMPEWAEDE 322
Db 280 TCG-RGELPYVTLKCC-----TAG-----VPTPYVPOYLRMPEWAEDE 322
Qy 339 CRKRRKLEENLQKQKDKSDSEYRYSRNGYDCEQTSIRKGRVGMKGCSTDFPAGSY 398
Db 339 CRKRRKLEENLQKQKDKSDSEYRYSRNGYDCEQTSIRKGRVGMKGCSTDFPAGSY 398
Qy 323 CRKRRKLEENLQKQKDKSDSEYRYSRNGYDCEQTSIRKGRVGMKGCSTDFPAGSY 398
Db 323 CRKRRKLEENLQKQKDKSDSEYRYSRNGYDCEQTSIRKGRVGMKGCSTDFPAGSY 398
Qy 399 ENWIDNQRQKQK--KYTEISDGGGR--KRAVGTTKYKYEKSYEKLKNDGYTVD 456
Db 399 ENWIDNQRQKQK--KYTEISDGGGR--KRAVGTTKYKYEKSYEKLKNDGYTVD 456
Qy 382 AEWIDNQRQKQK--KYTEISDGGGR--KRAVGTTKYKYEKSYEKLKNDGYTVD 456
Db 382 AEWIDNQRQKQK--KYTEISDGGGR--KRAVGTTKYKYEKSYEKLKNDGYTVD 456
Qy 457 AFLGLLNNKACKQITGKTKNEKVSNGGAGVGGSGGSGASGNTDEK--GFTYRSEY 515
Db 457 AFLGLLNNKACKQITGKTKNEKVSNGGAGVGGSGGSGASGNTDEK--GFTYRSEY 515
Qy 441 SFLOLSKEKICE-----RIQVGEETANYG-----NFENSDTSFHTY 480
Db 441 SFLOLSKEKICE-----RIQVGEETANYG-----NFENSDTSFHTY 480
```


Db 960 SPRLRLRFLRPLNKKRKEQK--AQHGAATGLOLRPGTVTDDSDRDPOTQLKRGINPDRFLR 1017

Qy 1046 QMEYTLGDIYRDLIFSGSDTTSVSKOTPSSNDNLKNVLLASGTFDEQERKNKYKEIK 1105

Db 1018 QMEYTLGDIYRDLIFSGSDTTSVSKOTPSSNDNLKNVLLASGTFDEQERKNKYKEIK 1060

Qy 1106 NFRKCSFRS-APRLVLS-----HP-----QTMWENKGYIMHGMICALSKDIAK 1150

Db 1061 GFLKQQTIVTSPSPDTSRTTYPVHPQTSYEKTPQQTWENGNPHTIMGMICALTYEDSGAI 1120

Qy 1151 GVEKKPKQIENPENLAMDANKRPKPOYQYTNVKKLDENSGTSRTTQTOQASD--NTP 1207

Db 1121 G--QPPQKVEDADKVLKTL--KPTANGIKWYLKEDNTSSAPMTSSSSSSGSDNPTTP- 1175

Qy 1208 TLTHFVRRPTTYFRNFEEKSGFCRKRRLKLOIKYDKCYKENG DYR-----CSGDGE 1259

Db 1176 KLEEFVILPTFFRFLHMGQNFCKRMRLKOIYKCEKGENGYGGRKOKTPOCSGYGE 1235

Qy 1260 AC-DSITHDYSTVPFSCPGCGKCHSSYRKWIERRKIEFHGOSNAYGOQT-DATRNNG 1317

Db 1236 DCEQDLSKYSTDYADLECPKCAHHCNRYKKWIEKKDDEFTEDERKPPROKDYVYGNKK 1295

Qy 1318 NTFDEKFCSTLETPDAKFLERLKNBPCK--TNRKYGDDIDFEKDSKTFOTEXCGPC 1375

Db 1296 GGGDNGFCITLKSLSDAQFLKTL--GSCCKDNSEDNNDKLNFSQPNETFVPATNCKPC 1353

Qy 1376 PKRKTNG-QNGNC--GVSGLNGCNDGDKSIDAKELAKMSSTTDYVWRVSDNTNFEQD 1432

Db 1354 SEKRIDCKENKCKNNGGGGTETCNGTTFITSENFKQKQTAKEFYWRVSDNNPNNGF-D 1411

Qy 1433 DLKDACQHANIFKGIKRDQWKGCVYGVADICEQTNINERTDG-----KEYIQIALF 1484

Db 1412 DLNFAQONAGFEKRIKRDWECGKVCYEVG---IPEKNGVYTSGENNDQIITITRGLV 1467

Qy 1485 KRNVNFLLEDYNTKINDKISHCITKKGESKJNGCEKNSKLEKWKIAEWENIKRRFN 1544

Db 1468 AHVQVNFLEDYNTKIKHISHCNSSEGYTCI---KN-CVEQWISTKRTEWNIKILN 1521

Qy 1545 DQYENKQOPRYNVKSLIELPKIAVNDQDQVTKL-----VFENSKGC--TLISNQNN 1598

Db 1522 EQY--KNPDTNVNTTLQDLOSQI---DFNKAIKPCGTLTKFEDSCJNGASSESKNG 1575

Qy 1599 KENDATIDCMILKLGKLVKAKNC--PGKPSGEKQSDCKEPPRLDDEEQON--PEENTLEPPK 1653

Db 1576 HEYDAIDCMILRLQDKIDDCKNHNAQNGENQAKC-EKISAPDEDDDEALTEENPVYQPN 1634

Qy 1654 FCEPTTQI-PBEKCGEFCGNKEBKDEKKEES-EEPAEESGPAEEPA-PTAEESET-- 1708

Db 1635 ICKRPPKPAKEKKGCEPAEKKEKEKEKTEVNTVAKPTEKEAADDPAGPADSEENPE 1694

Qy 1709 ETNPEPPGSGAPAPPTPAPPTPPPL 1738

Db 1695 EEKAPPE--EVEYTKKDAKAPVKTTPASPPPL 1722

RESULT 13

T28157

erythrocyte membrane protein 1 - malaria parasite (Plasmodium falciparum)

C.Species: Plasmodium falciparum

C.Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000

C.Accession: T28157

R.Yang, S.Q.; Wollish, W.S.; Gut, J.; Wu, J.; Ahn, J.; Petersen, C.; Fujioaka, H.; Alkawas

submitted to the EMBL data library, July 1995

A.Description: The molecular cloning and DNA sequence analysis of Plasmodium falciparum

A.Reference number: Z20479

A.Accession: T28157

A.Status: preliminary; translated from GB/EMBL/DBJ

A.Molecule type: DNA

A.Residues: 1-2212 <YAN>

C.Cross-references: EMBL:U01083; NID:91517813; PID:91517814; PIDN:AA06961.1

C.Genetics:

A:Gene: EMP1

Query Match	23.7%	Score 2863:	DB 2:	Length 2212:
Best Local Similarity	31.8%	Pred. No. 3e-122:		
Matches 757:	Conservative 245:	Mismatches 533:	Indels 846:	Gaps 88:
QY	6	GGSGQDEDAKIVLDFEQKVHDEYHGEAKNVSELSGLSLASILGETAF--TVKSMQT	63	
DB	5	GRRGT--MSAKVLEIDIGETIOKKAHSADPFRSLQKN-----FGEAKFYNGEINQP	57	
QY	64	ESKYTEL--LEAN--SKRNPCKKDGGANDYDRSVKEQAGYDNKKKAKCSGM-----	111	
DB	58	NSKLCELDHETIDTNTDGHNSPC-----BGRQTVRRP-----DNRQCTCKNRITKDSVD	106	
QY	112	----TCAPRRRLATCNKPNPNKNSN--DSKAKHLLALEVCAAKYEEESIKTHPKDS	165	
DB	107	MSVGCAPRRRLHLCSHNLESIGTNNYSSAKAKHMLLEVCYAAKFEEESIVKNTLEQLGH	166	
QY	166	KYPSGDFPMCTMLASFADIGDIINGRDLYLGNKKKKONGKETEREKLEQKLEIKFIH	225	
DB	167	H--TTEGICITAFASFADIGDIINGKDLILNPQ-----ESARRKQLEDNLIRKIFEDIY	218	
QY	226	DNL-----KQDEAKRYANGDEDPHYFLREDMWYANREYWGAMTCSKELDNSYFPAT	279	
DB	219	KELTSSRNKGTNGAERYK--DESGNYVYFLREDMWYANRLDIKAAITC--KAPGNAOYFENT	276	
QY	280	CNDTQSGPSQTHNKCRCCKDKGANKGKPRAGDDGTIVPTFFDYVPOYLRFMEFEAAEDFC	339	
DB	277	CSN--GEKPTGEKQCI-----DG--TYPTNLIDYVPOYLRFMEFEAAEDEC	317	
QY	340	RKKKKKLENLEKQCGKDKSDERYCRSNGDCEOTISRKGKVRNGKGCSTCFKCSYE	399	
DB	318	RRRLNLKATKNGGMDDEKCKEYCSRNGYDCTITISIKYSNRRCTCYLYCDDYV	377	
QY	400	NIIDNORQFOFKK-----YTKFISGGGKKRAVGGTTPYEG--YEKSFYKLNKNGY	452	
DB	378	KYIDKKKKEFEQKKCKENEYIRNNESSQNSK-----NYNMVETDFGNLKRK--Y	428	
QY	453	GTVDAPFLGLNNEKAKCDITD--GKIFNKEVNSGGVVGSGGSGTSGASGTIDENKSTFY	511	
DB	429	OSMNFELKLNSEPTCTNIIDEKRIDE-----TKDPEE--TFS	465	
QY	512	REYCORPDCGVOKGGGQWER-----KTVKKMYRSLKYKPIKNG--MVLLTSYK	562	
DB	466	HEYCDPCCWGLCKADQGYKRLYENDPQCIKK--YERKGEPEIDVLYTGK	520	
QY	563	VKDMMLIKMKKECCLTONSSDGSVGSVVTYTGASGSGSEKKELYDEKSYKHNEDQVYN	622	
DB	521	EKKDIIY--KLRECKT--DGNTGF-----KNE--EWNCY-----YQVG	553	
QY	623	VQGEVEEDDELKAGGGLCILRPKKNKEVSEAKSONNHADIQTFHDFRYUWVANHMKD	682	
DB	554	NDKCYLENGEELGG-----EKKV-----KQUDFLFWVANHMKD	588	
QY	683	SIHMYTKRLKSCI--SDGKTYMKSRGNCKKCDCEFEKMYVOKETEMKPIIDNHEGTGIDE-	740	
DB	589	SIEMSK--LSNOLCKDKKT--CIKKCNQCKCYEKMIGKKVQEWQIKKHDPKQDPRQGW	645	
QY	741	GYYTTLLELILKLOFLKEDTEBENTENSIDAEBAEFLKLOKILLENENNLAVUNAGTEQ	800	
DB	646	GRYF--VLEFVL-----BGDFETDIYKAUGDAREIIVHIOEMLOKKEQVULH--EDASNM	696	
QY	801	KTYMKLNLHNLNDATKC-----KDCP-----LPE	825	
DB	697	KTIIDELDLHELKAKCICIVNHKDNCCPDULSDSEDEDEDIPRONPCAKPSGSAHRLV	756	
QY	826	-----	825	
DB	757	NKVASNNHHKKKRLQVNVGVSSKLKGDAKGBYRKSGTIIKLOKICSTLNHNSNAKRGHT	816	
QY	826	-----	825	
DB	817	DQPCRKQSKSEMFTEDEGMKPAFISTKYVDIYMPRPROHCTSNLEYLOTNNKLNGN	876	

```
OY 826 ----- 825
DB 877 DINGNPNIINDSLGDLVLEAANYEADFIKKMYKKONDYKDNATICTRAMKYSFADLGDIIK 936
OY 826 -----EDKSRGSAD----- 835
DB 937 GTDMDNDNSGSEKTRDKLREIFDIIKKKHPIKEIYKEDTPYTKLREDWMEANRRKKIWEA 996
OY 836 ---PSP-----DIFIP----- 843
DB 997 MOCPTPNCSFPCKSYHTPLDDYIPQRLRMWTEMAEMFCKEOKROYGELYVACNGCKDECK 1056
OY 844 ----- 843
DB 1057 VCTNKSOCSTSCOMACENYKFNINMWKEQWDMKEIKYLLIQAOTTAANGPDTYSGLV 1116
OY 844 ----- 843
DB 1117 DENEPVYNLEFELYKENGKIGNPRDPPRAKRSKRETAAPASVAKNDVYSTAAGVYHQM 1176
OY 844 -----RPEEKEDNEDDEDEYR 862
DB 1177 GPHMCKTQTECEKTEDEOYENMYTFKNPPQYKDACICNTPRPCKESRKRSEDSDE-- 1234
OY 863 DDETAKEF--TEGSAT-----DPTSLDVCPIVGVKYLTKDNESLQDACSLKY 908
DB 1235 --EEKVYKETEKEKATEDAVDTGPPPAKKEATTTLDVCPYAGVLTKE--LENNACPTYX 1290
OY 909 GGNNSRLGRCV-----TPSGEPTTSSDKKALCVPR 941
DB 1291 -GPKAPTMKCIPTREKTAATGSESSGNGALQRAKRAVESGSPVTS--SGSICIPRR 1347
OY 942 RRLRYIKKIYMAT-----KTESPQASGSEASTSGSTPPDSKEALLKAFVESALE 994
DB 1348 RRLRYIOKLHMDASGNTVVSQGAQPOG--TSSPSEKPTPSDK--LRTAFIOSAIE 1401
OY 995 TPELHRYKKEK-----KAVQEGAGHGLPRVEGSPRY-----DPEDKLKE--GKIPD 1041
DB 1402 TPELMDRYKKEKEIKKEKKEKVANGL--VPSLNGCPPOGVTGDSPOSKLQOGVYIPR 1458
OY 1042 GFLROMFYTLGDYRDLIFSGSNDTTSVSKDTPSSSNDNLKIVILLASGTEQDEREKMKY 1101
DB 1459 PFLROMFTLGDYADLIFF--GKMD---IYIDTKNGDKD-----IARE----- 1496
OY 1102 KEIKRRCSTERSAPNIVSHP-----QTMWNNKXYIMHGMVCAJLTSKDIANGVEKK 1155
DB 1497 KRIRD-----AIBRVLKNMDSQPSDEKQRTWMEQNGEHIWMGICALYKKEDEKTEPLK 1552
OY 1156 POKINPENLDEANKKRRPPQOYTNYKLDENSGTSPRTTTOASSDNT--PTTLTHFYK 1214
DB 1553 -ONEGLKSLMDEKNNKPRDOKYQYDKYKLDENSGTSPRT-----NDHVPTPLTNFIS 1605
OY 1215 RPTYRWFPEMDESEFCREKRRKRLKQIKVDCKYENGVDVGCSDGEGACOSISHDVSTVS 1274
DB 1606 RPPYRRIYEMCEITCREKRRKRLKIKYECMDEKQKCSGDEGECBEIRKODSTYVD 1665
OY 1275 FNCPCGKHCSSRYKRIERKKIEFKKOSNAYGOQKTDATRNNGNTFDKEFCCTLETPWA 1334
DB 1666 FYCPCGKCYCRFYKMIKKKEDEYKOKRAYNNOKTDARRNN--DNASFSTLDTCTTA 1722
OY 1335 AKFERLKNGPCKTNYKEGGDID-----FEKDSKTFQHTFYCGPCPKTKTYNCGN 1386
DB 1723 GDFLDTLKNKGPCK-----NDNVDSGENKKLIFDENGDTFYKTOYCGTCSLNGFCNGGD 1776
OY 1387 CGVSLGNGCNDK--SIDAKEIKMSSSTDVYRVASDNTNFEESDGLDACOHANI 1443
DB 1777 CNVR--TNTCNKSARNTTITADT--KNGCNSAEIMLVSD--DINGSGNDLEACKNANI 1833
OY 1444 FKGIKRDVWKGCVYGVADICEQTNINERTDKEYIQIRALFRWVENFLEDYNNKINDKIS 1503
DB 1834 FKGIKENKMKCYVFCPSDVCGGLKRNND--IDQNIILIRALFRWLEFYLDYNNKIRKLN 1892
OY 1504 HCIRKGECSKICNGEKNKCLEKWI EKKIAPMENIKRRFNDQYENKD--QDPYNNKSIIE 1562
```

```
DB 1893 PCINNGEKAICNG-----CYEOWINOKRTEMTNINKNFQNYGVDDVTEMSSRSFLY 1946
OY 1563 ELIPRIAVVNDQNVNIKLCVFENSGCTLLISMTQNNK---ENDAIQCM:KRLGVAKNCP 1619
DB 1947 DLRIQIAATIDGNHNGLVKLVSVKCNCGNNSONGKEEGEDLVCLLQKLEKRAECK 2006
OY 1620 GKP--SGEKQSDCKEPPPLPDEEDQ--NPEENTLPPKFCPT--TOPPEEKGETCGK 1673
DB 2007 DNPETSGIPQOCVSPNHIIEDEQPLEEENTVHPKICDDVLKTOPJPEEPGETCEES 2066
OY 1674 EEKDKKEESEEPKAEESGPAABEPAPTAESEETETNPPEPGTGPAPSTPAPPTPD 1733
DB 2067 PGPDTYKEEEEEE--KKEE-----KKEEPPGLP--PPTPPPAAPSTPIPPKRX 2112
OY 1734 TPPLP--RPQADEPPDSTLLQTTIPFGVALALGSIAPFLK 1772
DB 2113 PPOVEKNPWEHPAVIAPALVTSTLAMSVOIGPAALFYFLK 2153

RESULT 14
T28431
variant surface protein 1 homolog A4VAR - malaria parasite (Plasmodium falciparum) (f
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: T28431
R:Smith, J.D.; Kyes, S.; Craig, A.G.; Fagan, T.; Hudson-Taylor, D.; Miller, L.H.; Bar
Mol. Biochem. Parasitol. 97, 133-148, 1998
A:Title: Analysis of adhesive domains from the A4VAR plasmodium falciparum erythrocyt
A:Reference number: 220486; MUID: 99094502
A:Accession: T28431
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3026 <SMI>
A:Cross-references: EMBL:L42244; NID:g3540144; PID:g3540115; PIDN:AAD03351.1
C:Genetics:
A:Gene: var
A:Introns: 2906/3

Query Match 22.0%; Score 2668; DB 2; Length 3026;
Best local similarity 25.4%; Pred. No.3.2e-113;
Matches 812; Conservative 277; Mismatches 618; Indels 1450; Gaps 99;

OY 5 GSGSGTODEDAKHVLEDFEQYVDEHGBAKNYVELKSLSLASIIGETAFTVSMQTE 64
DB 8 GSSGSGEDKDAKHVLDSTIGEXYKKEKVAANYSSQLKGTLSNALFENP---KGQTE 63
OY 65 SK-----YTELLEANSKRNPKCKGKGNDVDRFSVKEQAGYDNFKKKS--NGMTCA 114
DB 64 NPKCKLYEYHTNVKRGHREHPCR--KGTE--KRFSDVGGECDNFKIKDSKNNGGACA 119
OY 115 PPRRLHICKKNPNMNSNDSSKAKHD--LLAEYCMAMAKYGESITKHYPYDSKYGSQDP 173
DB 120 PYRRLHICLRLNLEISALD--KINDDTLADVCLALHGGOSTIDQIPYIQQVYSSSP 177
OY 174 --MCTLARSPADIGDILRGDLYLGNKKKKKONGKRETEREKLEOKLI:ELFKRIHNLKDK 231
DB 178 SOICIMLARSPADIGDILRGDLYLGNKKKKK-----LEKNLI:KIFGKIYEKLDLP 229
OY 232 EAQKRYNGDEDPNRYKLRDMWTANREYVWGMTCSSKEIDNSYPR:VTCNDTGQGPSQTH 291
DB 230 RAKDHYKDEPDNPNFQLRDMWNNANROEWKAITCHAG--ESDKYFR:VPAAGTGCTG--TQ 285
OY 292 NRCRDQKDGANAGPKAGDGVTTVPYEDYVPOYLRMFEEWAD:CRKRRKKLEMLEK 351
DB 286 GPKRCNDK-----KP-----GSNTDPPYTFDYVPOYLRMFEEWAD:CRKRRKKLEMLEK 336
OY 352 QCRGKDKSDERYTCRNGYDEBOTISRKGVBMGKCTDCFPACGS:ENMIDNOKRQDPK 411
DB 337 KCRGQDGSKEKEYCDPFGNDCGFTASGKHLYLMDNACACCFSCSD:FRKWLAKQKDEFEK 396
OY 412 QK--KTTKEISDGGGKRRKRAVGGTTYK----EGYEKSFYEKLKNDGYJYVDAFLGLLNNK 466
```


Oy 1731 ----- 1730
Db 2473 RRRKCFNDLKNIQENEVDNSLLEKIQHVAKNEGIDILKKNLPDONAISEICDAMKY 2532
Oy 1731 ----- 1730
Db 2533 SFADGDIIRGRSKIDPTNNKIEKELQIKFQIODNANSLKMLPELREKMMANRKE 2592
Oy 1731 ----- 1737
Db 2593 VWNAMTCVAPNDADHLKKKNNPGNKSQIIASQTEQTKCSHSEPPDYDIPEERYLQIE 2652
Oy 1738 ----- 1744
Db 2653 WSEYCKALKREKNDKMDCKSCIKSGATCEKEDEKCKECNDCKCKEYKNIWDKQSEF 2712
Oy 1745 ----- 1744
Db 2713 DQGNOLYKLYTQDRTGHPSTARARNPIETQKLEDSGNDPYSAKYLIDISTHCTDYKFS 2772
Oy 1745 ----- 1744
Db 2773 ETDRESNVAESPYPKYDKENCKCVNTPTSNNDKPSPLGSPFLPKPKMKFYPKIG 2832
Oy 1745 ----- 1758
Db 2833 IGVLPFTNMVADPIYTHETVAKTFPNNAVPOPHINPDKTVAAPPTKNILNEVLPSPAIYVG 2892
Oy 1759 VALAGSIATFLFKKTKRASYGNLFQILQIPKSDYDITPLKSNRYIPVSDRYKGR--TY 1817
Db 2893 IALAGSIATFLFKKTKRHPV-DLFSVINIRKSDYDITPLKSPNRYIPTSGKRYKNGTF 2951
Oy 1818 IYMGDSDEDKYA-FMSPTTDTVTSESEYEELINDIYVPGSPKXYTLEVVLEP---S 1872
Db 2952 TLKEIVGTDSGTYDHSIDT--SSSESEYEELINDIYHVLNKTILEVLEPSSKLS 3009
Oy 1873 GNNITASGKNTPSDTRN 1889
Db 3010 GNTIPTSGKNTPSDTRN 3026
RESULT 15
114602
variant-specific surface protein - malaria parasite (Plasmodium falciparum) (fragments)
C:Species: Plasmodium falciparum
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jun-2000
C:Accession: 114602
R:Yoss, T.S.; Felger, I.; Weiss, N.; Beck, H.P.
submitted to the EMBL Data Library, February 1998
A:Description: Identification of a conserved 5' flanking region of Plasmodium falciparum
A:Reference number: 218158
A:Accession: 114602
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 12135 <YOS>
A:Cross-references: EMBL:AF050740; NID:g2944094; PID:g2944095; PIDN:AAC05220.1
C:Genetics:
A:Gene: varph17

Query Match 20.8%; Score 2518; DB 2; Length 2135;
Best Local Similarity 29.4%; Pred. No. 1.3e-106;
Matches 676; Conservative 273; Mismatches 523; Indels 824; Gaps 82;

Oy 1 MATSG--SGGTODEDAHVLDIEFGOKVHDEYHGAKNVSELKSLAS-----IIGE 53
Db 1 MAPEDDPGGGKDDAKHMDPRIGTIOKOVHNAAEFRNLQRLSATINIKLIG- 59
Oy 54 TATVKSNOTESKYTEL---EANSKRNPKCKDKGNDVDFSVKEQAGYDNKTKKCSN 109
Db 60 ---NIETCSLEYKKYKHNGGADVSDKRYPKELSR-KYVDFSDKIGGQCTNEKMRBG 115

Oy 110 GMTCAPRRRLHCNKNFPNMANSN--DSSKAKHDLAEYCMAAKYEGES:KTHYPKYDSKY 167
Db 116 IGACAPYRRLHCHNLNLETIETNNYESNNAKNHLVDYCMAAKYEGDS:KTYTTHQHKY 175
Oy 168 PESDFPMCTMLARSPADIGDIIRGDLVIGNKKKQNGKETEREKLQI(LKEIFPKRIDN 227
Db 176 --DSQLCTVLARSPADIGDIYRGKDLFYGMTQKE-----KREDLEN LMDIFKIHSG 228
Oy 228 LKDKAOKRYNGDEDPNTYKLRBDMWTANRETVWGCAMTCSKELDNSTPRATCNDTGGCP 287
Db 229 LIG-GVKARYK--DTNTEYELREDMWTANRETVWMAITCDA---HGTYFRATCGDNISP 282
Oy 288 SOTHNKCRCDKCGANNAKPRAGDGVYTYPPYFYVQYLRWPEEMEDPCRKKKKLE 347
Db 283 SMAKNNCRCOKKDGPPDO-----VPTFYFYVQFLRWPEEMEDPCRKKKKLK 332
Oy 348 NLEKOCRGKDRSDERYRCSRNQYDCEQTIISRKGVNMGKGTGCFPACGSYENMIDNQR 407
Db 333 DALQCRKREKHGAKKLYDLDLNNYDCEQTAGKHDFEEDDCGCGOYSCARFVMTIDNQK 392
Oy 408 QPDKO-KKTYTEISDGGGKKRAVGG--TTYEGYEKSFYEKLKNDGICTYDAFLGLLN 464
Db 393 EFLKOKKKYGTETIS-LKSNKKRDAGISTKYVDGEKFEYELKSE-YRTVGEFLGLLN 450
Oy 465 EKACKDIDG--GKINFEVNSGGGVVGGSGTSGAGTNDENKGTLYRSFYCQPCPDG 522
Db 451 EYTCKEVRDDEKGTIDFTVNSG-----SASG-DEVNK-TYRTYTCACAPWC 496
Oy 523 GV-QHKGQNOWERK-----TKYKMRWSKLYRPINKMVL:KSLKVVNDMIL 570
Db 497 GAEOGRNCGVMKAKNDRCSPGNDYTKYK-----KELPIJGDKTKSEIY-- 542
Oy 571 KKNKKEFLT--QNSDSGVSV-----YTTGASGNSGKRELYDEKCY-- 613
Db 543 -ERYRKFCKNNGKANREGGVGSENGAASNDATTVGCGGNSJSLCEKTKCYXY 601
Oy 614 ---KNEVQKVN--VQGEVEEDDELKAGAGCILPNKKNKEVSEASQNNHAIQOTF 668
Db 602 KKENNDGKDNIFCVOGAMQ-----NSKKDKV-----RSY 632
Oy 669 HDEFYVVAHMLKDSIHMTKRLKSCIS--DGTKMKCRNGCKKCDPEKVVOKETEMK 726
Db 633 NAFPDWYHDLIDISIKRNEHNG-CINKDNKT--CIKGSCKDCFLKVVQOKEREMK 689
Oy 727 PIKDFKTOEGIEPGY---FT---TLELIL-KLQFLK--EDTEBNTENSIDAEAEEL 776
Db 690 LILHEHNTQGGFDKGEHQRIGFTHDVVLNVLDDKKEKLKIEGTYGIR-----EET 740
Oy 777 KHLQKIL-KLENNLAVYVNAAGTBEQKTLMDKILNHELDAITCKKCI:LPDE----- 826
Db 741 KHKEMLDKEKDAAGTGVASGTGPKNIMDKLIEHLEQAKCKCKQ)PEQOSLGRSLNPH 800
Oy 827 -----JRSR----- 830
Db 801 VYDDGSPKKRDKKRTNPCIYDITTEYAVLAGVNAQFQGEVRAKMLJRSRKNGETKSLE 860
Oy 831 -----GRSA----- 834
Db 861 GDIKKQFKNGRSGSELNDGICKIDNKYSNDIRGSTAGPCTGKDGJNBRFNAGTKMEGD 920
Oy 835 ----- 834
Db 921 NEVSATHRNLVYIPRROHCTSULEKLDLVYSKSNVNSDFLGDVLLAANEAQRTKH 980
Oy 835 ----- 834
Db 981 FAHKRDDHGIACRSVYSPADLADIIIRGRDMMDKDDGAKQKMDIKFKTIFGNLYESTLPK 1040
Oy 835 -----DPGP-DIFTP----- 843
Db 1041 GKYYDDQRTPOYKQLRBDMWEANRDQVWKAWECEKDIKCEDPJPVDDYIPQRLRMWT 1100
Oy 844 ----- 843

Db 1101 EWAEWYCKVQSEYDELLKCKGCKIKGVGCTSGSDCTPCAECTTYGQIKIPWEDQ 1160
QY 844 ----- 843
Db 1161 WNNMLLOTLLYQWAEETARYGTRAVSGDVGDKDPVQVFLLELOKONSGKTTYNTAAG 1220
QY 844 ----- RPE----- 847
Db 1221 YIHGEARYGECEVQKPCNTNGNQDKYFRREKPRDHDEACKCKDRPOQSAGGAGARSLP 1280
QY 848 -----KEDDENEDDED-----EVRDEDETAKEETEGS--ATD 878
Db 1281 SPRPVSDDDHSSDEDEDEDEDGGAEDENDEPASEEYKDDTEDEVEETAVSQPAAP 1340
QY 879 TTTSLDY---CPYGVKYLTKNESLQDACSLKYGGNNSRLGRCV----- 920
Db 1341 TTTTPGYTPACEIYKDLFEKPKNTFKEACTOKYGGNNSRLGKCTPTSGDKAATRGSGD 1400
QY 921 -----TPSGEPTTSSDKNGAICVPPRRRLYIKK 949
Db 1401 TTQNDSESGSEGSQHRAKHTSDASEKSAKSGEPT-----GGSICIPRRRRLYVYG 1455
QY 950 IVDWATKTESQASGSASSTSGST---TPDSKEALLKAFVESALETFFLWHRKKEE 1005
Db 1456 LTKWA-----EIOSQSQALSGQTPAGTPSOADPLAFAVESAAVETFFLMDRYKKL 1509
QY 1006 KKAVAQEGAGHLP-RVEEGS--PEYDEPDLKESGKIPDGLROMFYTLGDYRDIPLPSGS 1062
Db 1510 NADPOGSSLGGAAPLQNLGALGSEETPEYSLKSGTIPPDFLRLMFTYTLGDYRDIYGVK 1569
QY 1063 NDTTSVSKDTPSSSNDNLKNIVLASGSTEOEREMKMKYKEI-KNFRKCTERSAPNLVS 1121
Db 1570 QDVTKALE---ASGDN-----KSSKNPMQELSSKIEILKN-----GCTPPPTPYT 1612
QY 1122 H-----POTWENNGKYITHGAVCALTSKDKIAKGVKKPKIENPENLMDANKPK 1174
Db 1613 HSPSSGTTSSSMWKTNGOHIKGMICALTYKE---SGEKKIEQVKATDN--TDLFEKIK 1666
QY 1175 PROQYNTNVLKIDENSGTSPRTTQOASSDNRP--TTLTHFVKRPTYPFWMPEMGESPCRE 1232
Db 1667 DKYSYDQKVEIKENETBAKQPQ---DGLTPOTLLSNEFKRPYFRYLEEWGQNFCKT 1722
QY 1233 RKKRLKQIKYDCK-VENGDVGRCSGDEACDSISTHYSTVPFNCPCGCKHCSYRKMI 1291
Db 1723 RKRMLKDIITYECRNSDNPBGHCGDGHONNELKN-NMFGIDCRDCKECHKYKKW 1781
QY 1292 ERKIEFHKOSNAYGOOKTDATRNNGTF---DKPECKTLETWPDAAKFLERLKNPC 1346
Db 1782 HKKEDEFHONKNGYGEHEKLT--NGDNYSGGDNTNFCCQIKKKTAEDEFLKALRH--C 1837
QY 1347 KT-----NKEYGGDDIDPEKDKTFQHTYECGPCPKFTNCQNGNG-----VSGLNGN 1395
Db 1838 KNSDDTJKSEDEKKNKINDEKPNTEFNSTYCKACPIYGVTCNRGTGRCRPIKINEKN 1897
QY 1396 CDGDKSIDAKEIAKMSSTVDVVARVSDNDNTNTEGDDLKDACOHANIFKIRKDVWKC 1455
Db 1898 IEGBQ-----IDINILIDDGATND-D-NELHEKCKEYGLYTNLKKQEWKCQ 1942
QY 1456 YVCGVDICE-OTNIN-ERTDGEKEYIQIRALEKRAVENFLEDYKNINDKISHCIRKGGSK 1513
Db 1943 HKNSYKCELOKPLNSEYD--ERIPKILFERMIIDFIQYNNKSKEIRTC--TNDVNS 1998
QY 1514 CINCCEKNSKCLEWIKETIAEMENIKKRPNDYENKDP--DIVNKSILEE-LIPKTA-- 1569
Db 1999 CKECGKNCDCVEEMLKKSAMEKIKDYNOHFDVERIASRTKSFEBOGTFSYAKK 2058
QY 1570 ---VVNODDNVIRL--CVFENSKGCTLISTNQNKNKENDAITDCKLGVAKKNCPGKPSG 1624
Db 2059 AQEYVEKREDEKIMGCTGSNTDG-----QTEBENGDFITNLIDRLQTKIQTQTOHMK 2113
QY 1625 EKQSDCKEPPPLPDEE 1640
::: | | | | |

Db 2114 TORNSC--HPPPNDEE 2128

Search completed: May 1, 2001, 13:06:18
Job time: 251 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2001, 13:02:07 ; Search time 52.89 Seconds

(without alignments) 2408.011 Million cell updates/sec

File: US-09-508-967-1

Perfect score: 12100

Sequence: 1 MATSGSGCGQDDAKHVLD.....VNKKKEFEDEYPISDIWN1 2228

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_0401.*
1: /SID56/gcgdata/geneseq/geneseqp/AA1980.DAT.*
2: /SID56/gcgdata/geneseq/geneseqp/AA1981.DAT.*
3: /SID56/gcgdata/geneseq/geneseqp/AA1982.DAT.*
4: /SID56/gcgdata/geneseq/geneseqp/AA1983.DAT.*
5: /SID56/gcgdata/geneseq/geneseqp/AA1984.DAT.*
6: /SID56/gcgdata/geneseq/geneseqp/AA1985.DAT.*
7: /SID56/gcgdata/geneseq/geneseqp/AA1986.DAT.*
8: /SID56/gcgdata/geneseq/geneseqp/AA1987.DAT.*
9: /SID56/gcgdata/geneseq/geneseqp/AA1988.DAT.*
10: /SID56/gcgdata/geneseq/geneseqp/AA1989.DAT.*
11: /SID56/gcgdata/geneseq/geneseqp/AA1990.DAT.*
12: /SID56/gcgdata/geneseq/geneseqp/AA1991.DAT.*
13: /SID56/gcgdata/geneseq/geneseqp/AA1992.DAT.*
14: /SID56/gcgdata/geneseq/geneseqp/AA1993.DAT.*
15: /SID56/gcgdata/geneseq/geneseqp/AA1994.DAT.*
16: /SID56/gcgdata/geneseq/geneseqp/AA1995.DAT.*
17: /SID56/gcgdata/geneseq/geneseqp/AA1996.DAT.*
18: /SID56/gcgdata/geneseq/geneseqp/AA1997.DAT.*
19: /SID56/gcgdata/geneseq/geneseqp/AA1998.DAT.*
20: /SID56/gcgdata/geneseq/geneseqp/AA1999.DAT.*
21: /SID56/gcgdata/geneseq/geneseqp/AA2000.DAT.*
22: /SID56/gcgdata/geneseq/geneseqp/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	12100	100.0	2228	20	W93944
2	5074	41.9	2197	21	B18352
3	4817	39.8	2182	18	W22476
4	4817	39.8	2182	21	W77906
5	4802	39.7	3060	18	W22475
6	4802	39.7	3060	17	W77905
7	3928.5	32.5	2913	17	W00384
8	3197.5	26.4	2703	16	R70236
9	3197.5	26.4	2710	18	W22482
10	3197.5	26.4	2710	21	W77904
11	2929.5	24.2	1726	17	W00385

12	2497	20.6	1700	21	B18144	Plasmodium falcipa
13	1385	11.4	700	16	R70235	P. falciparum EBL-
14	1385	11.4	700	16	W22481	Plasmodium ebl-2.
15	1385	11.4	700	21	W77903	P. falciparum ebl-
16	1319.5	10.9	431	21	B18350	Plasmodium falcipa
17	782	6.5	440	21	B18146	Plasmodium falcipa
18	702	5.8	921	21	W22480	Plasmodium E31a.
19	702	5.8	921	21	W77902	P. falciparum ebl-
20	697	5.8	793	16	R70234	P. falciparum E31a
21	428.5	3.5	1604	16	R70105	TNF-R-EBA 175 fusi
22	425	3.5	1786	14	R41043	CD4-EBR175 fusion
23	418	3.5	1435	16	R70232	P. falciparum SAMP
24	418	3.5	1435	18	W22477	Silic acid bindin
25	418	3.5	1435	21	W77900	P. falciparum SAMP
26	406	3.4	248	21	B18151	Plasmodium falcipa
27	379.5	3.1	362	21	W77912	Plasmodium DBL gen
28	360.5	3.0	411	21	W77913	Plasmodium DBL gen
29	345.5	2.9	749	16	R70233	P. falciparum EBL-
30	345.5	2.9	749	21	W22479	Plasmodium ebl-1.
31	345.5	2.9	749	21	W77901	P. falciparum ebl-
32	325.5	2.7	1245	16	R70106	TNF-R-PI. viVax Du
33	321	2.7	2441	21	B18161	Plasmodium vivax D
34	315	2.6	1028	14	R41044	Plasmodium vivax D
35	315	2.6	1115	12	R13457	Duffy receptor. P
36	315	2.6	1115	18	W22478	Duffy antigen bind
37	315	2.6	1115	21	W77899	P. vivax DABP bind
38	314	2.6	1061	16	R70231	P. vivax DABP. PI
39	311.5	2.6	1663	15	R46608	Plasmodium falcipa
40	310.5	2.6	1979	21	B18171	Plasmodium falcipa
41	287	2.4	1308	21	B18167	Plasmodium falcipa
42	287	2.4	1588	15	R46605	Malarial PfEMP3 ep
43	272.5	2.3	2907	21	W57452	Human transcrip
44	271	2.2	2485	21	B18172	Plasmodium falcipa
45	265.5	2.2	1639	19	W54145	P. falciparum synt

ALIGNMENTS

RESULT 1	
ID W93944	W93944 standard; protein: 2228 AA.
XX W93944;	
XX	
AC	30-JUN-1999 (first entry)
XX	
DT	
XX	
DE	P. falciparum PfEMP1 protein.
XX	
KW	Erythrocyte membrane protein: EMP, PfEMP1; malaria; antioocclusion;
KW	glycosaminoglycan-like moiety; antiaggregational; antimalarial;
KW	antigen receptor; infected erythrocyte; rosette formation; blood cell;
KW	capillary occlusion; cerebral malaria; treatment; vaccine; detection;
KW	medicament; parasite; diagnosis; drug screening.
XX	
OS	Plasmodium falciparum.
XX	
PN	W09915557-A1.
XX	
PD	01-APR-1999.
XX	
PF	18-SEP-1998; 98MO-SE01675.
XX	
PR	19-SEP-1997; 97SE-0003386.
XX	
PA	(KARO-) KAROLINSKA INNOVATIONS AB.
XX	
PI	Barragan A, Carlson J, Fernandez V, Qijun C, Whilgren M;
XX	
DR	WPI; 1999-254692/21.
XX	
PT	New isolated malaria polypeptides
XX	

PS Claim 4; Page 67-74; 80pp; English.

XX This invention describes a novel Plasmodium falciparum erythrocyte
CC membrane protein (EMP), PfEMP1, which is capable of binding to a
CC carbohydrate which exhibits at least one negatively charged
CC glycosamino-glycan (GAG)-like moiety and has antiaggregational,
CC anticoagulant and antimalarial activity. The carbohydrates of the
CC invention are capable of acting as receptors for malaria antigens
CC present on the surfaces of malaria infected erythrocytes, by binding
CC to these antigens the carbohydrates prevent rosette formation by the
CC blood cells, this prevents occlusion of capillaries as is seen in
CC cerebral malaria caused by Plasmodium falciparum. The products of the
CC invention can be used to treat malaria or to vaccinate against it, or
CC used to design a model to identify compounds that bind to PfEMP1. The
CC carbohydrates, polyphosphates and antibodies of the invention can be used
CC as a medicament for dissolving the rosettes formed by erythrocytes
CC infected by a malaria parasite. The products can also be used for
CC detection, diagnosis and drug screening.

SQ Sequence 2228 AA;

Query Match	100.0%	Score 12100;	DB 20;	Length 2228;
Best Local Similarly	100.0%;	Pred. No. 0;		
Matches 2228; Conservative	0;	Mismatches	0;	Gaps 0;

QY	1	MATSGSGGTODEAKVHLDGFGKCHYDEYHNGEAKNYSVELKSGLSLSTLGETAPTYS	60
Db	1	matsgsgsgtgdedakhvldefgkvndeyhgeaknyvelkslslstlgetatfvs	60
QY	61	MOTSSTYTELLEANSKRNPCKKDGKNDVDFSVKEDAGVDNKKMSQNMTCAPPRHL	120
Db	61	motsktyelleanskrnpckkdgkndvdfsvkeagaydnkkmksqnmctcapfrrlh	120
QY	121	LCNKNFPMNMSNDSSKAKHDLALVCVCAAAAYEGESITHTYPKYDSKPGSDPFMCITLAR	180
Db	121	lcnknfpmmsndsskakhdlalveemaakyegaslthtypkyskpgsdpfmcitlar	180
QY	181	SFADIGDIIIGRDLVLCNKKKKKONGKETBRBKEQLKEIFPKKHIDNLKDEAKRYNGD	240
Db	181	sfaDIGdIIIGrDlYlcnkkkkngketerekleqkIkeIfkKhIdnlJkDeaqkryngd	240
QY	241	EDPNFYKIREDMWFAANETWGAATCSKEILDNSYFPAATONDRPGGSGOTHNKCRCDKX	300
Db	241	edpnfYkIredwfaanetwgaatcskeildnsyIfraTondrpggsqthnkcrcdckx	300
QY	301	GANAGKPKAGGDVTVIPTEFDYVPOYLRFMEWAEDFCRRKKKLELNLKOCRGKXSD	360
Db	301	ganagkpkaGdvTvlpytfdyppylrfveeaaedfcrrkkklienlekqcrGkxsd	360
QY	361	EYRYCSRNGYDCBOTISRKGVKRMGKCTDCCFACSGSYENWIDNRKQFDPKORKYTKEIS	420
Db	361	eyrycsrnyGceqtlbrtkgvrmgkctdcfcfacsyemwIdnqrkfDkqkYtkEIs	420
QY	421	DGGRRKRAVGCGTKKYGSEKSEFEKLLKNGYGVDAFLELLNNKRCRKITTOGGKRNK	480
Db	421	dggrrkrravGctkkygseksEfeKllknGygvdaFlEllnnEkrCkIttoGgkrnk	480
QY	481	EVNNGGCVGGSAGTSGASTNDENKGFERYSEYCOPCDCGVQHOHGNGOMEKTYVK	540
Db	481	evnnggcvvggsagtgasgndenkgftrYseyCpCpdcgvqHqngmEkrtyvk	540
QY	541	MRMSKLYKPIKGMWLLSLKLVKKDMMLLKMMKEPCLTQNSDGSVGSVVTGASGN	600
Db	541	mrwsklykpingkmwllslsklvkkdmmlkmmkecltqnsdgsvgsvvtgasgn	600
QY	601	SEKELYDENKCYRHNERYOVNVOGEVEEDDELKAGGLCILPNPKKNKEVSEAKSQNN	660
Db	601	sekelydenkcyrhneryovnvogeveeddelkagggilcIlpnPkknEvsEaksqnn	660
QY	720	HADIOKFNHDFEYVWVWVHMLKDSITHMTRTKLSCISDGRMKRNGCNKRCDFEKKVQ	720
Db	720	hadioKfthdFYvWvwmLkdsIthwctkrlksdsdgkEmkcrngcnkKcdcfekvWq	720

Qy	721	KETEMKPIKHOEIPQOEIPGYYFTTLEILIKQFKEDETEENTENSJDAEEAEELKHQ	780
Dp	721	ketemkpiqhftqgeipeyyfttleililkgfikedteentensidaeeaelkhq	780
Qy	781	KILKLENNNLAVNNACTEOKTLMDKLLNHELNDATKCKDCPLPEEDKSGRGSADSPDI	840
Dp	781	kilkleennnlavnnagteqltmdkllnlhelndatckdcplpeedksgrgsadspdi	840
Qy	841	FIRPEEKEDDENDDDEDEVRDEEFARKEETMGSTORTTSJLDCPIYGVKLTJDNDSL	900
Dp	841	firpeekeddenddedevedrdeefarkeetmgstorttsjldcpiygvkltjdnssl	900
Qy	901	QDACSLKYGGNNSRLGRCCTPGSEPTTSDDKNGALCVPRRRRLYIKKIVDMARKTESP	960
Dp	901	qdacslkyggnnsrlgrrctvptsepttsddkngalcvprrrrlyikkivdmarktesp	960
Qy	961	QASGEASSTISGSTTPPDSKEALLKAPVESAAIETFFLWHRYKEEKKAQAQAGACHLPR	1020
Dp	961	qasgeasstisgsttpdskaalikaivesaaetfflwhykeekkaqaqagachlpr	1020
Qy	1021	VEESPEDEPDKKEEKIDGFLROMFPTLGRJOLILSGSNDTTSYKOPSPSSNNL	1080
Dp	1021	veespeyopedkikegldgflromfptlgrjollilsgsndttsyskdpssnnl	1080
Qy	1081	KNIVLLASGSTEOEREKMNRYKEIKNFRCKSTERSAPNLVSHPTWMENNGRYIWHGWC	1140
Dp	1081	knivllasgstegerekmnykelknfrckstersapnlvshptwemngryiwhgwc	1140
Qy	1141	ALTSKDKIAAGVEKKPOKIEPNENLMDENAKRKPPOYOYTWKLIKDNESGTSPPRTTQOA	1200
Dp	1141	altskdkiaagvekkpokiepnenlmdenakrkppoyoytwklikdnsgtsprtqtqa	1200
Qy	1201	SSDTPPTTLTHFYKRPYFPMFPEEMESFPCREKRRKTLAKOIKYDCAYENDNVGRSGIDEA	1260
Dp	1201	ssdtppttlthfyrpyfpmfpeemefpcrerkrrktlakoikeydcayendnvgrsgideea	1260
Qy	1261	CDISTHDISTVSPFNCPCGGKSCSYRKWIENKKTLEFHKOSNAYGOQKTDATRRNGTFF	1320
Dp	1261	cdsisthdistvpsfncpgcgkscsyrkwielkktlefhknsnaygqktdatlrngtff	1320
Qy	1321	DKEECRTLETMPDPAKLEFLKNGPKCTKNKEYGDDIDEXOSKTFQOHEFYCGPCPKRT	1380
Dp	1321	dkeecrtletmpdakleflkngpkctknkeygddidexosktfqohiefygcgpcpkrt	1380
Qy	1381	NCONGNCVSGLNGNCDGKSIDAKETIAKRRSSTDVVRVSRDNDTNTFEGDGLDKACQH	1440
Dp	1381	ncngncvsglngncdgksidaketiakrssstdvvrvsrdndntntfegddlkdacqh	1440
Qy	1441	ANIFKGRKQVWVCGYVCVDIOEONINBRDQKRYIOIRALFKRWENFLFEDYKNIND	1500
Dp	1441	anifkgrkqvwvcgyvcvdioeoninbrdqkryioiralfkrwnflfedynind	1500
Qy	1501	KISHCIKRGESKNCINCENKSKLEKWIETKIAEENITKRRFNDOYENKQDPDYNVKSI	1560
Dp	1501	kishcikrgeskcncincenksklewietkiaeenitkrrfnodyenkqdpdynvksi	1560
Qy	1561	LEELIPIRTAVVNODNATIKLCVENENSGCTLISNTQNNKENAIDOMLKILGVAKKNCPG	1620
Dp	1561	leelipirtavvndnativiklcvnensgctlisntqnkenaidomlklilgvakncpg	1620
Qy	1621	KPSGEKSDCKEPPRLPDEEDONPENNTLEPPFCFPTTQPPEEKGEGTCKNKEEKKDEK	1680
Dp	1621	kpsgeksdckepprlpdeedonpenntleppfcfpttqppeekgetcnkeekkedk	1680
Qy	1681	KEESEBAKEESGPAEEPAETAESEETETNPEPPGTGAAPASTPAPPTPTPTPLRP	1740
Dp	1681	keeseepakeesgpaeeepaetaeseetennpeppgtgaapastpappptptptlrrp	1740
Qy	1741	QADPEPSTLIOTTIPFGVALAAGSTAFPLFKKTRKASVGNLFOLQIPIKSPDYDPIAKS	1800
Dp	1741	qadepfstliottipfgvalagstafplfktrkasvgnlfqlqipikspdydpirks	1800

QY	1801	SNKTIIPYSDKTKKKTITITMAGSDSEDEKRYAPMSTTIDVTSSESEYEELDINDIYVGSER	1860
Db	1801	snrylpyvsdykkykylymegeadedkyaifmsdtltvtseseeeyeeldindiyvgspr	1860
OY	1861	YKTLIEVLEESGNNTTASGKNPSPDRNDIONDGIIPSSKITPDNEMNOLKKEEISNMLQN	1920
Db	1861	YkTLIEVLEePsgnnttAsgknPsdRndiOnDgiPssKiTpdNeMnOLkKEEISnmlqn	1920
OY	1921	QPNPDVNDYTGNSSTMTNITTTSSRRHVVNDNNTNTMSRDMMEENLLPSIHDGNTLSGEE	1980
Db	1921	qpnvPndyTsgnsTnttIttSRhVndNntNtmsRdmEeNllPsIhdgnTlSgEE	1980
OY	1961	YSYNNVNVNSNDIPINBDNNVTSGLIDLINDSGSGKRPIDITYEVLKREKNELFTEENRK	2040
Db	1961	ysynnmvnmndiPirndnvnysgIdlIndsIsGkRpidIyEvLkRkeNelfTeenK	2040
OY	2041	RSTQNVAKTTNSPIRHNOLELPHKVLDRHRDCKEKKNEEDILINKLKEPMKEENINNSG	2100
Db	2041	rTsqnVakTtnsPiRhnQlelPhkVldRhrDckEkKnEdIlInkLkEpmKeeNiNnsG	2100
OY	2101	KTYNSDNKPSNNHVLNTDVSIOIDMNPRTKNEITNNDTNQDKSTMDTIDDDLEKYNDEP	2160
Db	2101	kTynsdnKpsnnhVlncdvsIqIdmnpRtkneItNndtnQdksTndtIddleKynDp	2160
OY	2161	YYDFEEDDIHYHDVVEKSSMDIYVDHNNVTSNNMDVPTKKMHIEMNIYNNKKEIFEEX	2220
Db	2161	yYdfEeddiYhDvVeKssmdIyVdhNnVtSnnmDvPtKkMhIEmniYnnKkeIfEeX	2220
OY	2221	PISDIWNI 2228	
Db	2221	pIsdiwnI 2228	
RESULT	2		
ID	B18352		
AC	B18352	standard; Protein; 2197 AA.	
DT	07-NOV-2000	(first entry)	
DE		Plasmodium falciparum chromosome 2 related protein SEQ ID NO:210.	
KW		Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;	
KW		antimalarial; malaria; protozoacide; infection; insecticide.	
OS		Plasmodium falciparum.	
PN		MO200025728-A2.	
PD		11-MAY-2000.	
PE	05-NOV-1999;	99MO-US26796.	
PR	05-NOV-1998;	98US-0107131.	
PA	(HOEFF/) HOFFMAN S.		
PA	(CARU/) CARUCCI D.		
PA	(GARD/) GARDNER M.		
PA	(VENT/) VENTER J C.		
PI	Hoffman S, Carucci D, Gardner M, Venter JC;		
DR	WPI: 2000-365347/31.		
XX		Proteins encoded by chromosome 2 of the human malarial parasite,	
PT		Plasmodium falciparum, useful as antimalarial vaccines and in the	
PT		diagnosis of P.falciparum infection -	
PS		Disclosure: Page 441-447; 577bp; English.	
CC		The present invention describes proteins and their fragments (I) encoded	
CC		by chromosome 2 of the human malarial parasite, Plasmodium falciparum.	

Also described are: (I) nucleotide sequences (II) encoding (I); and (2) vaccines against *P. falciparum* infection comprising (I) or (II). (I) and (II) are useful for the development of vaccines against *P. falciparum* infection. (I) and polyclonal antisera or a monoclonal antibody raised to immunogens comprising the sequences of (I), are useful in the detection of infection with *P. falciparum*. Furthermore, (I) (especially when they are rifins or secreted or membrane proteins) can aid the identification of drugs to treat or prevent *P. falciparum* infection, or they can be used to identify drug resistance in *P. falciparum*. Sequencing of the Plasmodium chromosome 2 and the subsequent identification of proteins encoded by it will help to expand our understanding of parasite biology, a process hampered by the complexity of the parasitic lifecycle, and provide new targets for vaccine and drug development. Parasite resistance to drugs and mosquito resistance to insecticides have led to a resurgence of malaria in many parts of the world, and there is a pressing need for vaccines and new drugs. A70078 to A70287 and B18144 to B18352 represent nucleotide and protein sequences given in the present invention, but which are not specifically mentioned within the specification.

Query Match	41.9%;	Score 5074;	DB 21;	Length 2197;
Best Local Similarity	47.1%;	Pred. No. 6.9e-298;		
Matches 1134;	Conservative 277;	Mismatches 599;	Intels 400;	Gaps 877;

```
0Y      4 SGGSGCDDAEKAHVLDGFGQVHDEYVGEAKNVYSLEKSLASTLJG--ETATVSMQ   62
          | | | | | | | : | : | : | | | | | | | | | | | | | : | :
Db     3 gkggdpdesvkxlmfdrlgedvyeqsvselnyvsleqklsiapllgyesstclcn   62
          | : | : | | | | | : | | | | | | | | | | | | | | |
0Y      63 -TESYTYTELIEANSKRNMCKCKDKGNDNDRFSVEOAGYDKKRKNCSGM?-----C 113
          | : | : | | | | | | | | | | | | | | | | | | | | |
Db     63 lvgdylnmpygnsnrlypcnkljgtneerlstdtlggscnnkkikngystksqkdccac 122
          | : | : | | | | | | | | | | | | | | | | | | | | |
0Y      114 ADFRLHLCKNFPMNSNDSSKAKHDLLAEVCMAATEGESIKTHNPKTDSKTPGSDFP 173
          | | | | | | | : | : | : | | | | | | | | | | | | | | |
Db     123 apyrxlhcsh---nlesldtsmtbhklllevcmaakyesnidthypqhrtnebpsq 179
          | : | : | | | | | | | | | | | | | | | | | | | | |
0Y      174 MCTMLARSPADIGDIIRRDRIYLGRKKKKONGKEPERKLLOXKEIPIKKIHDLXORE- 223
          | | | | | | | | | | | | | | | | | : | : | | | | | | | | |
Db     180 lctmlarstfadgdvlrvykdllyngskeke----kdeletnlkt.fgklnexlkkeq 233
          | | | | | | | | | | | | | | | | | | | | | | | | | | |
0Y      233 AQRYNGCEDPEFYELRLDMWTANRELYWAGMATSCEKLEDNSYFRA.CNDPGQPSOTHN 292
          | : | : | : | : | | | | | | | | | | | : | | | | | : | :
Db     235 aeTry-gsdctnyglredwyanaratweatc--dvhgsdyfry.cgdktetatlvkd 299
          | : | : | | | | | | | | | | | | | | | | | | | | |
0Y      293 KCRCDKDGANAGRKPAKGDDVTIPTYEDVYPOYLRFMEEMADPF.RKKKKKLENLEKQ 352
          | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     292 kercdengkkpg-----snadqpyflydvvpqylrwfeaeadf.rkkkkkklekleg 345
          | | | | | | | | | | | | | | | | | | | | | | | | | | |
0Y      353 CEGKRXKSDEYR---YCSNRNGIDCEBTISRRKRVNMGKCTCFPACZSYEWIMDNOKOF 408
          | : | : | | | | | | | | | | | | | : | : | : | : | : |
Db     346 cr-----dyknhlycsgnnygdctkltykkbklylgehtclncswcslyeswlndqglef 399
          | : | : | | | | | | | | | | | | | | | | | | | | |
0Y      410 DKOK-KYTKEISDGS-----GRKKRAYG--GNTYKEGYEKSPYEKLKNDGCGTV 455
          | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     400 lkqkkyetelstnsngscggsggvkgrnkktkgayetelnldgyekkykelksegykv 455
          | | | | | | | | | | | | | | | | | | | | | | | | | | |
0Y      456 DAFLGLINNEKACKDITD-GSKINFKEVNSGGGVYGSGSGETSGASGTENENKCTFYRSE 514
          | | | | | | | | | | | | | | | | | : | : | | | | | | |
Db     460 odffklilmedvccklkdekchidfk-----pdchnsmegtfhshe 502
          | : | : | | | | | | | | | | | | | | | | | | | | |
0Y      515 YCOPRPDCGVONHGKGNOMERKTAYKMMMSKLXPINKMMYLLKLKYVADMILIKNNW 574
          | : | | | | | | | | | | | | | | | | | | | | | | | : | :
Db     503 yckprdcpsvark-dnwykkdydgactr-qklylpasgaqtpblkljksekkqketkl 560
          | : | : | | | | | | | | | | | | | | | | | | | | |
0Y      575 KEFCLTONSDDSVGSVYTTG---ASCENSEKKELEYDEMKSCHNIEVOKV-----NQV 624
          | : | : | : | : | : | | | | | | | | | | | | | | | |
Db     561 kafc---dqtnqdtlnsarvaggaagsgsksnknelyleewkcy--neygkvdkdkgnee 615
          | : | : | : | : | : | | | | | | | | | | | | | | | |
0Y      625 GEVEDDELKAGAGCLTLPPKKNREVSEAKSONNHADLOKTEHHFFYYVAWHMLDSTI 684
          | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Db 616 dedeadvkvkaggjcllen---khneshnassnepegfgktfnfffwjgrflndsm 672
 QY 685 HMRTRLRKSCISDGRKTMCRNGCNKKCDCCFEKWKXOKETEKKPIKDHKTQE----- 736
 Db 673 ywrgk-vnscimnpkirkkrnecckddcgclfwgkkeeuenlkkhktlqeaflnkren 731
 QY 737 -----GIPRGYVFTTLELILKLOFLKEDTEENTENSLDAEAEELKHLKIL-----K 784
 Db 732 sgldmfsglmds-advvleleleleqflqdlkgd-----ygdvkelqkelldeekkk 784
 QY 785 LENENNLAVNVAAGTQKTLMDKLNHLENDATKCDCLPEEDKSRGASDPEDIFIP- 843
 Db 785 kgaeeavvvvadnqkktldkllqhegdanncl-----kthkeceetpkp-----pg 835
 QY 844 -----RPEEK-----EDDENEDDEDEVRDEDEETAK--ETTEG----- 874
 Db 836 agggapapetcttlededeedeedeagadeveegelevdtegedetevqpykdtre 895
 QY 875 -----SATDTTSLDVCPIVGKVLTKDNESLDACSLKYG-GNNSRL-GMRCVTPSGE 925
 Db 896 geeeaakaldttsldvcdtlnalt-nqnlldacklkypgpkkerfpmkcvcs-sge 953
 QY 926 -----FTSSDKNGAICVPRRRRLYT---KTLVMAITTESQASGEASSTSG 972
 Db 954 kavalagsagaigksgdgaicvprrrrlyvggllkllsagtsesppg-9gsessrsasd 1012
 QY 973 -----STPPDSKEALLKAFVESAALETFFLMHRYKE-----KKAVAQE-----G 1013
 Db 1013 vsqnggddtltt-----eslrkwlfeicaaletfflmhrykweaqqkaelqnglllg 1067
 QY 1014 AGHGLPRVEEGSPEDPEDKL-KEGKIIPDGLRQMFYTLGDRILFSGSDTTSVSKDT 1072
 Db 1068 tgaal---nlqgdpsnptqigkagclpdlflmflyclgyrdllvgvad----- 1116
 QY 1073 PSSNDNKNKTVLLASGTSEDEBEKMKYKE-INKNRKCS---TERSRPNLVSHPQTWME 1128
 Db 1117 ---dknggnmlilnasgn-kdekqmeklqelqllpfsngketrgpnvndrqsllwd 1172
 QY 1129 NNGRYIWHGMVCAITLSK--DKIAGVEKKPQKIENTPENLMEANKRPPOYQYTNVLD 1186
 Db 1173 rlaehvwhgmccaltkxddnglkgyvkkpkqlenpeklnetckkpkdekyqgqtaile 1232
 QY 1187 ENSGTSPTTQTOASSDNTPTTLTHFVKRPYFRWFEEWGESFCRERKKRLKQIRVDCKV 1246
 Db 1233 desg-----ekrpsasagklclfdlkrprrfrylgeengfckrtlemigklkede-y 1285
 QY 1247 ENGQVRCSGDGEACDSTSTHDYGTVPFNCPCGCHCSSRKRTERKKTIEFHQSNAYG 1306
 Db 1286 knq--gtrsgdglkcnelivldkellfgollpcrcahrfrykwlntkrdrdefnkysnays 1343
 QY 1307 QOK-----TDATRNNGNTFDEKFECKTLETPDAKFLERLKNGPCKTNKE---YGGDD 1356
 Db 1344 eqkkkyeendsaqknq-----vcgltk--ddaaeflnrlknpgckneeseenkkaede 1395
 QY 1357 IDEFKDSTFQHTREYCGPCPFKTNCGNGCVSG-LNGNCDGKSIDAKELIARMSSTT 1415
 Db 1396 idfkppddtfdadncpkceefklkcnhncsngntgkcdgkttlaateienlknltk 1455
 QY 1416 DVNVRVSDNONTNTEFGDDLKACQHANTEKIRKDYMKCGVYGVDC---EQGININERT 1472
 Db 1456 evtlnlvddksatefcdgjlseckdkgllfgllrkdwecsgkvgcvdclnllkdhnlqhes 1515
 QY 1473 DGRFYIRALFRRMVENFLFDYNNKINDKISHCIKKGEGSCINGCEKNSKCLERKIEKK 1532
 Db 1516 d-kkylilmkellktwlefyiledynklkhklncknngyskclg-----cvdkwvqk 1568
 QY 1533 IAEWENIKKRPNDQYENKNDQDY--NVKSILEELIPKIAVNDQDNVILKLCVFEENKSGCTL 1591
 Db 1569 keekqglkerfneqykaktseyfnvksfletwlpklavndqgnvklksfgnscgsa 1628
 QY 1592 ISNTQNNKENDADICMLKCLGVKAKNCBGRKPSGKOSDCKE---PPPLPEED--QNPE 1645
 Db 1629 saistngneadalcmiklekideckkrkpgensgqcnctllhpldvgedepleeete 1688

QY 1646 ENTL--BPPKFCPPPTQPPPEKGGGTGKNKEKKDEKKEESEBPAKESSPAEAPPTA 1703
 Db 1689 enpyvgkqnpstcpvpeakkkkeegetc-----tpasapa--- 1723
 QY 1704 ESEETETNFBPPPGTGAAPSTPAPPTPPPLRQADEPFDSTILQTTIPRGVALAL 1763
 Db 1724 -----papapaspapcpap-----adeqfclrlqtllpigtialal 1759
 QY 1764 GSIAFLPK-----KTKASVGNLQPIILOIPKSDVDIPRLKSSNR 1803
 Db 1760 gsiaflflkvlvlycvvyymylymcfcilymkktkhrv-dlfsvlnlpxsdydlprklspr 1818
 QY 1804 YIPYSDRKGTGYIYMGDDEDEKAFMSDPTDVT--SSESEYELINDIYVGSPPRYK 1862
 Db 1819 yipytsqkyrgkryilylegdsqts-gytdhysdltssesesejeemadndlyvpspkyk 1877
 QY 1863 TLEIVLLEPSSGNNTTASGKNTPSDTRNDIONDGISSKITDNEWNOLEKEFISMKLONOP 1922
 Db 1878 tlievnllepagnnta-----sdtqndiqndgipnskfadnewntlkddflsmllgnqr 1931
 QY 1923 NDVPRNDITSGNSSTRTNTITTSKRNVDNNTWTYSRDNMEENLLPSTHDGNLYSGEYS 1982
 Db 1932 kdvpndyksgdipfnctq-----pntlyfdkpeekpfltslhdnllngeys 1978
 QY 1983 YVNMN-VNSMNDIPLNRDNVYSGIDLINDSLSGKRPIDYDEVLKRRKENLFGTEWTKR 2041
 Db 1979 yvnmnschsmdd-pkyvsnnyvsigdlindslsngkhdlydevlkkkenelfctmnhvk 2037
 QY 2042 TSTQNVAKTNSDIPINQOLEFHKWLDRHDMCEKWKNNKEDILNKLKEWNNKINNSGK 2101
 Db 2038 tsihsvaktndsdplnqinlfhtwldhrmcckewhberlakikeew--eneethsn 2095
 QY 2102 TYNSDNKPSHNHVLNTVYSIOIDMDNPKTKNEILNMDTNDOKSMMDIILDLEK-YNDPY 2160
 Db 2096 thpsds-----nklntdvsqilmdnprkplnqfn-----mdtllledlkrpnepy 2142
 QY 2161 YUDEVEDDIYHDVDEKSSMDIYVDH--NNVTSNNNDVPTKNHIEINYNKKEIFEE 2218
 Db 2143 yudmyddd-lyydvn-----dhdstvtolnamdypskvqlend-vntk--lvke 2187
 QY 2219 EYPISDIWNIT 2228
 Db 2188 kypiladvwdi 2197
 RESULT 3
 W2476
 ID W2476 standard; Protein; 2182 AA.
 XX
 AC W2476;
 XX
 DT 12-SEP-1997 (first entry)
 XX
 DE Plasmodium var-1.
 XX
 KW DBL gene family; SABP: stalic acid binding protein; vaccine; therapy; Duffy binding like gene; Duffy antigen binding protein; erythrocyte; DMBP; merozoite; malaria; var-1; var-2; var-3; var-7; Immune response; Plasmodium.
 KW
 OS Plasmodium vivax.
 OS Plasmodium falci-parum.
 XX
 FN W09640766-A2.
 XX
 PD 19-DEC-1996.
 XX
 PF 07-JUN-1996; 96WO-US09508.
 XX
 PR 07-JUN-1995; 95US-0487826.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

xx Chitnins C, Miller LH, Peterson DS, Sim KL, Su X;
PI Wellens TE;
xx WPI: 1997-052231/05.
DR N-PSDB; T72883.
xx New malaria vaccines - contains cysteine-rich DBL family protein
PT binding domains homologous domains of the Duffy and sialic acid
PI binding proteins
xx
PS Claim 8; Page 70-74; 96pp; English.
CC This sequence represents var-1 of Plasmodium. Var-1 belongs to
CC the Duffy binding like (DBL) family of genes which have homology to the
CC Duffy antigen binding protein (DABP) and sialic acid binding protein
CC (SABP) conserved regions (see T72889 and T72888 respectively). The var
CC family of genes modulate cytoadherence and antigenic variation of
CC Plasmodium infected erythrocytes. SABP and the Duffy antigen binding
CC protein (DABP) are soluble proteins that appear in the culture
CC supernatant after infected erythrocytes release merozoites. DABP and SABP
CC mediate the binding of merozoites and schizonts to the erythrocyte
CC surface. These proteins are necessary for erythrocyte invasion by the
CC parasite. This sequence can be used in the compositions of malaria, and
CC comprise either a nucleotide sequence or encoded polypeptide of the
CC var-1, var-2, var-3 or var-7 genes of the DBL gene family, a family of
CC genes having homology with conserved regions of DABP and SABP. The
CC compositions are used for the treatment and prevention of malaria. They
CC are also used in the preparation of vaccines for inducing a protective
CC immune response in a mammal to Plasmodium merozoites (especially
CC Plasmodium falciparum or Plasmodium vivax).
xx
SQ Sequence 2182 AA;

Query Match 39.8%; Score 4817; DB 18; Length 2182;
Best Local Similarity 45.9%; Pred. No. 2.4e-282;
Matches 1096; Conservative 298; Mismatches 607; Indels 388; Gaps 93;

QY 4 SGGSG---GTOD-----EDAKHVLDEGQKVHDE--VHGKAKNVSELSKSLSLAS-I 50
DB 11 sgsssgsgkdktdseylyvsdskdlldrygekyeevkngdakkyialklnhtangr 70
QY 51 LGETAFAVSKMOTESKYTELLEANSKRNPCKKDGKNDVDRFS-----VKEQAG 99
DB 71 ssetasietctclvkeyervngdgkrlpcrdaknedvnrtsdclggqctynrlsdsg 130
QY 100 YDNKMKKSNMGTCAPFRRLHLCNKNFPMNSNDSSKAKHDLAEGYCAAKKEGSIKTH 159
DB 131 gdnkv-----gacapyrrllhcdy---nlesldtstthkllleavmaakkeygnslnth 181
QY 160 YPKYSKYPGSDPFCMTMLARSFADIGIIRGRDLYLG--NKKKONGKETREKLEOKL 217
DB 182 ytcqhrtnedessqclctvlasfadiqidvrgkdllygdnakey-----rkllqql 235
QY 218 KEIFRKIH-DNLKDEAKQKRYNGD-EDPNFYKLREDMWTANRETYWAGTCSKELDNSSY 275
DB 236 kdlfkkllhdkwmktngageryldakgqdfqlredwtsnretvkwallchapleanyf 295
QY 276 FRATONNDGGGSGSOTHNKCRCDDKAGANAGRPKAGDVTYPTTFDYVPOTLRMFEEMA 335
DB 296 lktacn-vykg---tngschc-----lsgd---vptyfdyppqylrtfwea 335
QY 336 EDFCKKKKKKLENLKRGKRGKSDERYCSRNGYDCBOTISRKGRVWRCQTCQFAC 395
DB 336 edfcgkkkkkllenlqgctdyegn---lycsngydccttklykklivgecltnscwvc 392
QY 366 GSYENMIDNQRQDKOK-KYTKEISDGGG-----RKKRAVGTTKYE--GYEKSFEYEL 447
DB 366 rmyetwidngkkelikgrykyeteisggsqspkrtkraarssssdngyeskfykkl 452
QY 448 KNDGCTYDAFLGLNLKACKADITDGGKINKKFNKFNVSNGGVSNGSGSGISGASGTNDENK 507

DB 453 keyrygdvdkflklklnkaglcqkqpyg--nekahn-----vdfnekyv 495
QY 508 GTFYRSEYCOQPCDGVQHKGGNOMERKTKYKMKRMSKLYRPNGKNVJLLKSLKRVKDM 567
DB 496 flfstrleicepwpwgle-kqgpwv--kvkgdklctgsaktlydknl:ldplvlypdksq 552
QY 568 MLTKNMKEFCLTQNSSDGVSQSVYTTGASGNSKSKELDYDMKCYKHDEYQKAVVQGEV 627
DB 553 gnllkkykxf-----ekgapggqjlk-----wqcy 580
QY 628 EEDDELGAGAGLCLTPPKKNKEVSEAK---SONNHADIQKTFHDFEYVVAHMLDS 683
DB 581 ----dehr-----pskmmnnevgtwdfitgkqt--vksjnyfvfwkvhmlnds 626
QY 684 IHWRTKRLKST---SDGKTMKCRNGCNKCKDCFEKWKQKETEKKPLKDHFKTQGEI-- 738
DB 627 vewkte-lskclnnhtnglctcrnnkckldcgcfgkwkvekkqemalikhgkqtdlvq 685
QY 739 -----PEGYFTTLEILK-----LQFLKEDTEENTENSIDAEALELKHLOKILKLE 786
DB 666 qkglivfipyg---vldvllkggnllgnlk-dvngdt-----cdlkhlkkl-ld 729
QY 787 NENNLAVVNAAGTEOKTMDKLNHELNATKCK---DCPLPEEDKSIGRSADPSPDIFI 842
DB 730 eedavavvlygkdntli-dkllqhkeqaeqckqgeceekkaqges/grissetredert 788
QY 843 PRP-----EKEEDDENEDDEDEVARDDETAKETTESATDTT'SLDV----- 885
DB 789 qgpadsageveeceededdddydeddddvveeegke--egvtevtvyleveevtveq 846
QY 886 -----CPYGVKVLTDNSLSDACSLK--GGNSRLGRCVTPSGEPTSSDKNGAICV 938
DB 847 egvkpodiavgk-lfeddkslkeacgllkyppgkexfpmkcvtpsgv:latsgkqgaicv 905
QY 939 PRRRRRLYIKKIVDMATKTESPQASGEASSTSGSTTPPDSK---EALKLAVESALTEP 995
DB 906 pprtrrlyvgglsgwastr-----gddetlevsseatsapseseekltaliesaalet 959
QY 996 FELMHRKYEKKKAVA-QEGAGHGLPRVEGSPEDYDEPKLKE-GKIPDGLRMEYTTIGD 1053
DB 960 felwmklyeeekprptqdgaglyvalpepppgedpqlqglylfpoffllgmlytlad 1019
QY 1054 YRDLIFSGSNDTT-SYSKDTPESSNDNLKNLYLVLASGSTEGOREKMN-----KYKEIKNFR 1108
DB 1020 ykdllysgsndtsdltgqtpssndnlknlvlaeysgstegkekmlqigaklklilnga 1079
QY 1109 KCTERSAPNLYSHR-QTWEMENNGKYIMHGWCALTSKDKITAKGYER-K-PQKIEPENIM 1166
DB 1080 tsqvpvptknsyvkcpqgqlwenaIdkdlvnamvcallykendargts:akegnkdlklalw 1139
QY 1167 DEANKKPRPQYQYNNVNLKDNSGTSPTTQOASSDNT--PTTLLIPEVKRPTYRWEEE 1224
DB 1140 deannntrplekyqynvnlkeds-----aksndltgprtlk:lfelptrflwhe 1190
QY 1225 WGESFCREKKRKLKQIKVDCKVENGDVGRSGDGAQDSISTHDS:VPSFNCPCGKHC 1284
DB 1191 vgnsficferakrlaqklhncmeddge-kqysgdgyceel:ifakqynv:lqdlas-scacapc 1248
QY 1285 SSYRKWIERKKTLEFHKOSNAYGOOKTD-----APFNKNNTDTE:FECKTLLEWPDNA 1335
DB 1249 rlykvwlekkkleyeqqkayeqksnyeneqkdcqqtssnna---wetsrl:lgasptaa 1306
QY 1336 KFLERLKNQPCCKTNEY--GGDD--IDFEKDSRTFOHTEYCGPPEKFTNQCNGNGVSGSL 1392
DB 1307 eflqll--gscndnngyengednkldfknprkdtfeahscdpcrptlykcgqhc-vgsa 1363
QY 1393 NG-NCDGDSIDAKETIAKMRSTTVVMRVSPNDNTNTEF--GDDLKDACQAHNITKGIK 1449
DB 1364 ngkecknuk-ltaedlkhktdpnugnliemvsdsntfeliqgd-----cksagllkglirk 1417
QY 1450 DVMKGYOCGVNIG--EONINERTDQKEYIOIRLFRKRWVNFLEDVKNIDKISHCIR 1507

Db 1418 dewkancvgydiclekklknqgegdkkyitmkellkrlwylefyedynlrkkklck 1477
QY 1508 KGESEKCIINGCEKSKSCLEKWEKKEIAEMENIKKRENDQYENKQOPDVNVAKSIIEELIPK 1567
Db 1478 kegdckckly-----clekvwgektkewqkndtyleqyknd-----gnltlnfleg 1525
QY 1568 IAAVNDQDNVILKLC-----VFENSKGCTLIISNTONKKNDAIDCMKLKLVAKNCPKPS 1623
Db 1526 fgyrtelfnaikpcgdjldqfctscglstnshsgnmndvlcllnklqkkliseckeghs 1585
QY 1624 GEKQSDC-----KEP-----PPLPDEEQNPENMTLEPRKCPPTTQPREKGE--TCG 1671
Db 1586 gqctcpncslsgksetlyevdvdyeqnp-ekkvgeqpkcpdmkepkendevegtcg 1644
QY 1672 NKEKKEKKEESEPAPKEESGPAEEPAEPAE--TAESEETETN-PPEPGTGAAPAPSTPA 1728
Db 1645 gdeekck--kvedsvleqkkeeaasapeespllpeapkkeenvvpkpp----- 1690
QY 1729 PPRDTPPP-----LRPADEP-FDSTIIQTTFPGVALAGSIAFLFKKTKASVGN 1781
Db 1691 -----pppkrrrlktrnvldhpavlpalmastlmwslgigfaaflylkkkckssvgn 1744
QY 1782 LFOILQTPKSYDTPTLKSSKRYTPYSDRKGYTYIMEDSD-EDKYAAMSOTTDTYS 1840
Db 1745 lfqlglpksqdydipcltkssnrylpyasdrhkgklylmeqdsqdekyatmsdtldts 1804
QY 1841 SESYEELIDINDIYVPGSPKKTLEVLVEPSGNNTTASGKNTPSDPTKNDIQND---GIP 1897
Db 1805 sesyeeldindiyvpspklyllevlps-----krdcqndlnhdipadip 1853
QY 1898 SS-----KITDNEMNQLKEPISNMLQNOPNDVPNDYTSGNSSTNTNTTTSRHVNDNTN 1953
Db 1854 nsdcppltdedewnglkkdfismnlqntqtepn-----lhhvndnth 1898
QY 1954 TTNSRDNMEEMLLPSTHDGDLVSGEESYVNV-----NMVN-----SMNDIPI 1996
Db 1899 pcmrthmdqgpfmsldhrlfsggeeyuymfngsnplisdstmsdlsfnhnsby 1958
QY 1997 NRDNVNVSGIDLINDLSGCKRPIDYDEVLEKREKNELEGTENFKRTSTQN--VAKTNSD 2054
Db 1959 ndkndlysgldlindalsgnh-idlydemlkrkenelfgcthpknltsnrvvqtsdd 2017
QY 2055 PIHQLELFHKWLDHRHDMCEKMKNKEDILNKLKEEMNKENINNSGKTYNSDNKPSHNHV 2114
Db 2018 pltnqlnlfhkwlgrhdmcekwknhberlpklkelw--enethsgdt--nsglpsghv 2073
QY 2115 LNTQVSTQIDMDNKTKNETNNMTNODKSTMDTILDDLEKYNDRPYDYDFEDDIYHDV 2174
Db 2074 lntvtslqldmndpklmefmtdmtpdkstmdtildlekynepyydyfkhk-lyydv 2132
QY 2175 DVEKSSMDIYVDHNNVTYSNNMADVPYTKMHIEMNIVNNKKEIFEDEYPS 2223
Db 2133 nddkasedhlmndhkmndnnsdvpntvqlemvlnn-qellqneypis 2180

RESULT 4
ID Y77906 standard: Protein; 2182 AA.
XX Y77906;
XX AC Y77906;
XX DT 13-JUN-2000 (first entry)
XX DE Plasmodium var-1 polypeptide.
KW DBL gene; Duffy-binding like gene; ebl-1; Duffy Antigen Binding Protein;
KW DABP; Stalic Acid Binding Protein; SABP; malaria; vaccine; immunisation;
KW protozoacide; var-1.
XX OS Plasmodium sp.
XX PN US5993827-A.

PD 30-NOV-1999.
XX
PF 07-JUN-1995; 95US-0487826.
XX
PR 10-SEP-1993; 93US-0119677.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Slim KM, Chitlins C, Peterson DS, Su X, Wellens TE, Miller LH;
XX WPI; 2000-194198/17.
DR N-PSDB; 298288.
XX
PT Isolated protein binding domains from Plasmodium vivax and Plasmodium
PT falciparum erythrocyte binding proteins useful for vaccinating against
PT malaria -
XX
PS Disclosure; Columns 129-140; 93pp; English.
XX
CC The invention relates to ebl-1 polypeptides that are encoded by the DBL
CC (Duffy-binding like) gene family. The ebl-1 proteins are substantially
CC identical to the Duffy Antigen Binding Protein (DABP) and Stalic Acid
CC Binding Protein (SABP), which are soluble proteins that appear in the
CC culture supernatant after erythrocytes infected with malaria release
CC merozoites. Immunochemical studies indicate that DABP and SABP are the
CC respective ligands for Plasmodium vivax and Plasmodium falciparum Duffy
CC and stalic acid receptors on erythrocytes. The ebl-1 polypeptides may be
CC used to vaccinate against malaria, especially caused by P. falciparum.
CC Immunization with the polypeptide provides effective protection against
CC malaria. The present sequence represents the var-1 polypeptide.
XX
SQ Sequence 2182 AA:

Query Match 39.88; Score 4817; DB 21; Length 2182;
Best Local Similarity 45.94; Pred. No. 2,4e-282;
Matches 1096; Conservative 298; Mismatches 607; Indels 388; Gaps 93;

QY 4 SGCGS-----GTOD-----EDAKHVLDPEFGOKVHDE--VHGEAKNYSELKSGSLAS-I 50
Db 11 sgsssgskgkddtegylyvsdaklldrvgekyeevknkgdakkyllealngnlntngr 70
QY 51 LGETAFTVKSQNTESKYTELLEANSKRNPCKKDKGNDVDRFS-----VKEOAG 99
Db 71 sseassietcltkeyeryernvgdgrchrckdaeknedvnrfdatlgsqctynrlksqg 130
QY 100 YDNKMKCSNMGTCAPRRRLHLCKNKPNNMNSNDSSAKHDLAEVCMAAKYBESIKTH 159
Db 131 gdnkv-----gacapyrlhldcy--nlesldtstcltklllevmaakyegnslnth 181
QY 160 YPKYDSKYVPGSDFMCMGLARSPADIGDIIRGRDLVYG--NKKKKONGKETEREKLOKL 217
Db 182 ytcqqrtnedsaagqlcvtlatstadlgdlvrgkdllygdnkeqg-----tklqgkl 235
QY 218 KEIFKKTH-DNLKDEAKQRYNGD-EDPNFYKLEEDMWTANRETVMGAMCSKELDNSSY 275
Db 236 kdlfkklhkdmktnqsgeryldakgddffqledwtsnretcvkallchapkeanyf 295
QY 276 FRATCNDTGCGPSQTHNKCRCDDKGANAGPKAGDGVTVIPYEDYVPOYLMEBEEMA 335
Db 296 lktacn-vgky---tnqgchc-----lqgd-----vptyfdyvpqylwfeewa 335
QY 336 EDPCKRKKKKLENTLEKQCRGDKSDERYCSRNQYDCEQRTISRKGKVMKGCIDCFAC 395
Db 336 edlcrkkkkkllenlqkqcrdyeqn---lycsngydccltkykgklvlghecnscwvc 392
QY 396 GSYENMIDNQRKQDKOK-RYTKELISDGG-----RKRRNVGTTKXE--GYEKSEFEXL 447
Db 393 rmyetwldnqkkelkqkrtyeteisgsgsgkspkrtkrraarsassddngyeakfykkl 452
QY 448 KNDGYTVDAFLGLNNKACKDITDGGKINFEVNSGGGVGSGGSGTSGASGTNDENK 507
Db 453 keyvgdvdkflklklnkeglcqkqpyg--nekadn-----vdflnekyv 495

Oy	508	GTFRSEYCOBPCDPCCVQHNGWGNOMERKTYVKMKRSKLTKXRNKGMLVLLSKLAVNDM	567
Db	496	kfirtelcepcwgcgle-kygppw--kvygdctgsaktikydpnkltidprlypsksq	552
Oy	568	MILKMKNKECFLTONSSDGSVSGSVTTATGASGCKEKKELYLEDMCKYKHNHEVOKVWOGEV	627
Db	553	gnlltkykntc-----ekgapgggqlkk-----wgcy-----	580
Oy	628	EEDDELKGAGGLCLIPNPKNKVESEAK---SONNHADIKTFFHFEFYWAHMLKS	683
Db	581	----dehr-----psknmnncvegwkdfitgkqt-vksyvnffwdvwmhlnds	626
Oy	684	IHMRTKRLKSCSI---SUGKTMKRCNNKCNOCEFKWYQOKETEMKPJIXDHKKOBEGI--	738
Db	627	vewkte-lskclmnnugnrcnmnckctdcgcfqkwkekqgemmalldhbgjktldvg	685
Oy	739	-----PGCYFTYTLELLK-----LOFLDETEENTNSIDAEBELEKHLOKLTLE	786
Db	686	qkgjlvespyg-----vidvlkgsnllpnlk-dvhgd-----ddiknklkl-Ld	729
Oy	787	NENNLAYVNAAGTEBKTLMDKLHLHELNDATKCR---DCPLPEBDKSKGRSADPSPDIPI	842
Db	730	eedavavnlqgkduttl-dklhqkeaqeqckqkqecekkaqgesirjsaetredert	788
Oy	843	PRP-----EEKDEDNEDEDEVRDEDETAKETEETEGSATDTPTSLDV-----	885
Db	789	qgpdsasgeveeeeddedydeddddvqgeeeyke--egltvetlevteveeetreq	846
Oy	886	-----CPYGVKVLTKDNESLDACSLKY--GGNSNRSLGRCVTPSGEPTTSDDKGAICV	938
Db	847	egvpkcdilvgk-lfiedkksikeacgilkypgggekfpmkcvrtpsgvatstsqkgaicv	905
Oy	939	PPRRRLRIKKIYDMARKTSPASSOSEASTSGSTTPDSK---FALLKAHVESALET	995
Db	906	prrtrrlvygglqsuar-----ggdetevseatsapsaseeklrtaltesaat	959
Oy	996	FFLMHARKEEKKAANA-OEGAGHGILPRVEEGSPREYDEDEKLK-GKIPODFLRQMVTGLD	1053
Db	960	fllmhkykeekppraqdgaglgvslprespdpdqtblqtlqvlpkoflrlqmtylad	1019
Oy	1054	YRDILEGSDNTT-SYSKDTPSSSDNLKNIYLLASGSTEOBEREXNN---KYKEIKNR	1108
Db	1020	ykdliygsndtstcltgkqtpssndlnkhivleaaqstegekemkqigakikklinga	1079
Oy	1109	KCTERSAPNLVSHR-OTWMENNKGXYIMHWALCSKOKIKGYEKK-POKITEBPENLM	1166
Db	1080	tsgvpyvtksvaktpqgtweenlakdtimawcalilykendargtsakeqnkdlkalw	1133
Oy	1167	DEANKKRKPROYOXTNVKLIDENSCTSPTQTQOASSDNT-PTTLTHFYKRPTRYRWEE	1224
Db	1140	deanknpriekeygtynvkledesg-----aksndltippltkfnfelptfirthe	1190
Oy	1225	MGSEFCGEERKKRKLOIKLVDCXVNGDVGRCSGEGECADSISINDYSTVPSENCPCGCNHC	1286
Db	1191	wgmfceteratragihnemcedage-kxysgdygeceefiskxyvnlqdis-saccapc	1248
Oy	1285	SSYRWMTERRKIEFHKOSSNAVGOOKTD-----ATRNNGTEDPECCKTLEFWPRAA	1335
Db	1249	rlytwtlekkkkteyekqkayegqkenyenegqdcbqsgnma--netstrilgsptrea	1306
Oy	1336	KFLERLANGCEKTINKEI-GGDD--IDPEKDSKTFQHTYECCGPCPKFTMCONGNCVSGL	1392
Db	1307	efiqkl--gsecindngengeendkifdknpdklfikaahscdpbltlyvcqgnhc-vysa	1363
Oy	1393	NG-MCDDDKSIDAKEIYKMMRSSTTDYVMRYMSRDDWTMFE--GDGLKDACOHANIRKGRK	1449
Db	1364	ngkechmkn-lteadinknctdpgnlienmvssdstntufelhgd-----cksqgikfglrk	1417
Oy	1450	DVMKCGIVYCGVDIC--BOTNINERTDOKEYIQIRALEFRMVENLEEDYWKINDKISHCIK	1507
Db	1418	dewicanvugcdltclekklnggedqkytkmlkelkylwilefyfedyorlfrkkklck	1477

QY	1508	KGESEKTCINGCEKSKSCLEKWEIKKEMIKEMENIKRRFENDQYENNDQDY VKS1LEELIPK	1567
Db	1478	kedgckcikg-----clekwgvekikewqkindylegyknd-----gnltfnlieq	1525
QY	1568	IAVVDNDQNVKILC-----VFENSKGCTILNFTDNKKNENDAIDCMKLIVKKNCEGRKS	1623
Db	1526	fgytetfknalipcdqldqfktscglnstlsgnqnmndvlci lkk lseckeqns	1587
QY	1624	GEKOSDC-----KEP---PPLUEDQNPENNTLEPPKFCPTTQYPEEKGE--TGC	1671
Db	1586	gqtqpcdnsslsqksteelvedvdyeeqnp-enkveqpkfcdmkeqkкендеevctcg	1644
QY	1672	NKEEKDKKESEEPKAEESGPAEERAP--TAESDETETN-FPEEICTGPAAPSTPA	1728
Db	1645	gdeekc--kvedsylegkeeeaaapeespl lpeapkkkenvvpkpf-----	1690
QY	1729	PPTPTPTPP-----LKRQDEP-FDSTILIQTLIPFGVALAGSINAFILKTKTKASGN	1781
Db	1691	-----pppkrriklctnrvlndhpavipalmsstlmslgiglaafylylkkktsvgn	1744
QY	1782	LFQILQIKRSDYDIPITLKSSNRYIPYSDRKKCTVITYMGDSO-BDIYAFMSDPTDYS	1844
Db	1745	lfgqlqpkpsqsydpltklssnrylpyrsdtrhkqktylymegdsqgeiyafmsdtdlts	1804
QY	1841	SESEYEELINDIVVPSPKYKTLIEVLEPPSGNNTTASGKNTSDTINDIOND---GIP	1897
Db	1805	seseyeeldindivvpspkyltievlepp-----krdtyndhndpsldp	1855
QY	1898	SS-----KTIDENKQLKKEFTSNMLQNPDPVDRDITSGNSSNTNTI TTSHKNDNTNT	1955
Db	1854	nsdlppitdewnglkkdfismalqntqtlepp-----llhndvndnth	1896
QY	1954	TTMSDNDNEENLLPSIHGDLVSGEYSYV-----NMV-----SMNDPI	1996
Db	1899	ptmsrhmndqpfmslndrnltsgeeynydmfnsqnpnlndstcnsmdsjltsmhsby	1956
QY	1997	NBDNVVSGIDLINDISLGGKPIDIYDEVLKRENEELFGTEKTRKSTON--VAKTNSD	2054
Db	1959	ndkndlysgidlndalsgnh-ldiyemlkrkenelfgqghphkhltenrvvtqtsad	2011
QY	2055	PIHNOLELPHKWLDRHRDMCEKKNKEDILNKLEEMKNENINNSGITYNSDNKPSHNHV	2114
Db	2018	pltnqinfhkwldrhrdmcekwnkherlpxkkelw--enehsgdl--nsglpsgnhv	2072
QY	2115	INTDVSIOIDMDNKTKNETNMDTNOCKSMDPTIIDDLKRYADPYVYFYEEDIDITHYD	2174
Db	2074	lntdvsiaqidmnpkctmefnmdntlmpdksctmtilldlekynepyydfdyknd-lyydv	2133
QY	2175	DVEKSSMDIDVVDHNNTVSNMMDVPTFMHLEMMNIVNNKKEIFEEXYIS 2223	
Db	2133	nddkasehlimdnhkndmnnstvdprvtqlemvinn-gelllqneyjls 2180	
RESULT	5		
ID	W22475		
XX	W22475	standard: Protein; 3060 AA.	
AC	XX		
XX	XX		
DT	XX	12-SEP-1997 (first entry)	
XX	XX		
DE	XX	Plasmodium var-7.	
XX	XX		
KM	XX	DBL gene family; SAMP; stalle acid binding protein; vaccine; therapy;	
KM	XX	Duffy binding like gene; Duffy antigen binding protein; erythrocyte;	
KM	XX	DABP; merozoite; malaria; var-1; var-2; var-3; var-7; immune response;	
XX	XX	Plasmodium.	
OS	XX	Plasmodium vivax.	
OS	XX	Plasmodium falciparum.	
XX	XX		
PN	XX	W09640766-A2.	

[illegible]

Dd	326	oKncrcrgkdkedkdyrcsrmnydcektkraiajkllygkqcsclyaenpydw4lnqke	385
Qy	408	QFDKO--KRYTKEI-----SDGGRKRAAGV--TTKYEGEKSEFYEKLNKGDTVDAP	458
Dd	386	qfdkqkkkyddeeikkyengasgsgrqcrtdaggttcttnynygdekfydelnkseyrtvdf	445
Qy	459	LGLLNKACKADTID--GGKINFEVNSGGVVGSGSGTSGASGTNDENKCTFYRSEXC	516
Dd	446	leklsneeicckvkxdeegatldfknv-----sdstsgsgynvsaggtfyfyskyc	496
Qy	517	OPCPCCGQV--HKGG--NOMERKTKVKMKRSLYRP---INXKMWLLLSKLKVVKDWMI	569
Dd	497	qpcpycygkkyvngsgssnweekm-qkcsqkilyepkpkdegttctllksqghd---	552
Qy	570	LKKMKREFCLYTONSSDSVGSVYTTGASGNSKKELIDEMKCYKHNEVOKNVOGEVE	629
Dd	553	leeklnkccdekngdtlnsgsgsgsgsgnsgqrgeyleemcykykedvkvynhdedee	612
Qy	630	DDDELKGAAGLICILPNPKKNEVSEAKSONNHADIOKTFHDEFYVVAHMLKDSIHWTK	689
Dd	613	dymcnknaagjicllknqkknkeegntesekepdelqgtfnffyyvalmlkdslnhw-k	671
Qy	690	RLKSOISDGKTMKC--RNGCNKCKDCPEFKWOKOKEETEMKPKIOHFKPODEIIPGYFTTLE	748
Dd	672	k1qclqgnaricqgmknkmdcecfirwltqkdwgklyvqnllqknikqrgysdntae	731
Qy	749	LI-----LKLQFLKED-----TEETWENSLDAEEAEELKHLQKILKLENNNLAV	793
Dd	732	lipfhdvlygnlqgeelkygdsdaseeksensldeeeekhlrelteednmgeas	791
Qy	794	VNAG--TEOKTLMDKLNLHELNDATCKDCPLPEEDK-----	828
Dd	792	vgyvteqknimdklnyekdeadlclelhedeekygdgnecleegenfrypcsgse	851
Qy	829	-----	828
Dd	852	sgnkrypvlankvaygmhbaktqlaasragrsalrgdislaqfkngrngstlkgqlckln	911
Qy	829	-----SRGRSADS-----PDIFRP-----	843
Dd	912	enysndsrngsgpccgctkgdghgvrmrlytewsnlegkqkcsyknvfliprrhemcsn	971
Qy	844	-----	843
Dd	972	lenldvsvrtnkdnkashsllygdvqlaaktadaaelkryrkqdmqnlqtdbqlqkdgeamcr	1031
Qy	844	-----	843
Dd	1032	avrysfaqlgdlfgrdmwmedkssctmetrlltvlfnlkekhgdlkdnkytgdskkpr	1091
Qy	844	-----	843
Dd	1092	aykklradwearnrhqyramkcatkyljicpmpvddyipqrlrwmtewaeayckagsqe	1151
Qy	844	-----	843
Dd	1152	ydklkkicademskygdcgkcgdvdcgkckaadcckykeelekwneqwrksdqkynlllylq	1211
Qy	844	-----	843
Dd	1212	aktstnpgrtvljddpdyqymvdfllpikhaslaarvlvkraagspelaaaapitpy	1271
Qy	844	-----	843
Dd	1272	staagylyhgeigygcqeqctqfceekkhgatssttkenkeyctfkpppeyatacdclms	1331
Qy	844	-----RPEEKED-----	850
Dd	1332	qteepkpkkeenevesacklveklllegkngrtvtvgcnpkhesypdwcdkmlidshdgacmp	1391
Qy	851	-----DEN-----EDDEDEYVRD-----	863
Dd	1392	prtgllcilylahesqtenlctddnllkdfatkaaaetflswgykksknsaaklildtql	1451

```

OY 864 -----DETFAKE-----TWEGS----- 875
Db 1452 lpsqflrmmytfgdyrdicIntdtskqndvakakdkiqkftskdgskspsqlstrqew 1511
OY 876 -----ATDP----- 879
Db 1512 ktngpelwkgmcaltkyvtddnkrkikndysdkvngsqngnpsleefaaqpqlrm 1571
OY 880 ----- 879
Db 1572 lewgeefaeqrqkenilkdaeneinstqgcndakhrccnacrayeryenkkkefsqgt 1631
OY 880 ----- 879
Db 1632 moflkanvpqdpkykyeykdqvplqneyllqkcdnkcscmdgnlsvspkexpf 1691
OY 880 -----TSLDVCPIYGVKLTDKDNEISLODA 903
Db 1692 gkyahkypeekcdcygqkhvpslpppppyvqgppeaprtvvdcsiv-ktlfdttnfsda 1750
OY 904 CSLKTYGNNLSRLGWCV---TPSGEPT---SSDKNGAICVPPRRRLYIKKIVDMATKT 957
Db 1751 cglky-gktapsswkcipstdksagatqgksdgsisciprrrrllyvgklgwatal 1809
OY 958 ESPQASGSEASTSGSTTPPDSKEALLKAFVSALETFFLHRRTRKEEKAAVQBGAGHG 1017
Db 1810 --pqegaaapshra-----ddlnafiqsaaletfllwydkyeeekbpqg-qgsqga 1858
OY 1018 LPRVEE--GSPYDEDEK-LKEKGIPOGFLROMFTYLGDRDILRSGSDMTSVSDRPS 1074
Db 1859 lsglscstysddeedppdkllqgkllpdpdlrmfyllqdyrillvhggn--tsdagnclg 1916
OY 1075 SSNDNLKNIIVILLASGSTQERENKMYKEI--KNFRKCSSTERSADNLVSHPOTWENNCK 1132
Db 1917 snnn---nlvleasgnekmdqklegqlpknngstlrvpssr---aqtpkwwnehae 1970
OY 1133 YIMHNMVCALT---SKDKIAGVEKKPOKIEPNENLME-----ANKR-----FKP 1175
Db 1971 slwkgymicaltyeknptlsargde---nklekdeveyekffgsstadvktagstcltgyk 2027
OY 1176 POYOTYNYKLDENSGSTPPTOTQASDNTPTTLTHFVFRPTYPFWFEEMGESFCERKK 1235
Db 2028 tqdyekykletdsg-----aktpsasestp-lisdflvtrpyfyleegwgnfckkrh 2081
OY 1236 RLKQIKVDCXVENGDVG-----RCSGDGACDSISTHDYSTVPSFNCPCGKRHSY 1287
Db 2082 klagklhckckveenggsstrgiltqysgdgacnemlpkndgtvrdlekpcakrcsey 2141
OY 1288 RKMIRKRIEERKQSNATGOOKTDTATRNNGNTFDKEFCITLETWIDAAFLERLKNPCK 1347
Db 2142 tkwieskglekqkqeqdk-dkcvnngsnkhndgfcetltskakfkl-tpck 2198
OY 1348 TTKREYGGDDIDEXKSKTFPOTREYCGPCFKTNCONGNGCGVSLNGCDGKSIDAKET 1407
Db 2199 punvqkctifd---ddkftkhkdcpcrkfsynckdecd-nsgytdcrnknsidardi 2254
OY 1408 AKMRSSSTDVYVRSVNDNTNFEQDLEKDACQHANIFKGIKRDVWKCQVVCVDICEQTN 1467
Db 2255 engvdstv-lemrvsadsksfgngdlenacrsgagifeglrtdewkcrnvcgvvyckpkn 2313
OY 1468 INERTDKEYIOIRALFKRWVNFLEDYKINDKISHCIRKKEGSGCINGCEKNSCLEK 1527
Db 2314 vngaeqkghlqiravtrrweyffedynklkhkshrlkngelspc1---kn--cvek 2367
OY 1528 WIEKRTIAEMENIKRRPNDOYEKRDQDVWKSILEELJPKIYVNDQDVAIKLVEPENS 1587
Db 2368 wtdqkrkewkelterfkdykndnsddanvrsfletllpqldanaknkvllskfsgnc 2427
OY 1588 GCTLISNTQ--NNKENDALDCKLKGVRKAKNCPK--PSEGEKOSDCKEPPPLPDE--- 1639
Db 2428 ggsaaneqngkngey/kdaidcmllkklkldigceckhhqbstctesdtpqpqlleedtid 2487

```

```

OY 1640 ---EDQNPENITLPEPKFCPP--TOPPEKKGEGTCGNKEKKDEKE:SEEPKAEESGP 1694
Db 2488 ddleceekaknmnm--pkicenvlktagedeg--c-----vpaeiseepatdsqk 2535
OY 1695 AAEPAPPAESESETETNPEPPGCGAPPSRPAP--PTPPTPPLR-3OADEPDSITL 1751
Db 2536 etpeqtrvlkpeeavpeppp-----ppqekapaplpqppppppptjldnphvltal 2589
OY 1752 QT-TIPGVALLGSLAFLELKKTRKASVGNLFQIOLIPKSDYDPTLKSNNRIPYVS 1810
Db 2590 vstlswsvyigfatfkyllkktksvgnlfqllqpkpsdydipklspryipytsq 2649
OY 1811 RYKGTITYMEGSDPEDKXAFMSDPTDVTSSESEBELINDIYVGSFKYTLVEVE 1870
Db 2650 kyrgkytlylegdsgtds--gycdhysdltseseyeeamdindiyvpskxytllevle 2708
OY 1871 P-----SGNNTTASGKNTPSDTRNDIONDGIIPSSKITDYIEMNOLKEEFISN 1916
Db 2709 psqmnttasgnttasgnttasgntktpsdctqndiqndgipsskitdiewnqlkdeifsq 2768
OY 1917 MLONDPNDVNDYTSGNSTNTNITTTSRHNVNDNTNTTMSKDNMBEILLPSIHGONLY 1976
Db 2769 yjqsepteqn-----mlgyvndntthpctshnvee:pflmsindrnlf 2813
OY 1977 SGEESYNY-----NMVN-----SMNDIPINDNNVYSGIILINDISGGRPI 2019
Db 2814 sgeeynydmfnsgnnpnlisdtsnmdsltsnhsbpndkndlysgljlindalsqn-h-1 2872
OY 2020 DIYDELRKENELFGE--NTRKTSQNVAKTNSDPINQOLEFELKALDRHRDMCEKK 2078
Db 2873 dlydenlkkkewkelfgtkhtkhtntnyvakparidpctnglnflfkxldrhrdmcekk 2932
OY 2079 NKEDIILKKEPMKNENINNSKTYNSDNKPSHNHVLNTDVSIOIDVNPRTKNEITNND 2138
Db 2933 nherjpklkely--enechsgdi--nsglpsgnhvlntdvsiqldnrdpklkneitnnd 2988
OY 2139 TNOCKSTMOTIIDLLEYKNDPYYPYEDDITVHDVDKSSMDITVYDHNVTSSNMNV 2198
Db 2989 tnpdsktmctlliddlekynepyydieddillyhdvdeksmddilyvhnvntnmndv 3048
OY 2199 PTKMHTENIVN 2210
Db 3049 ptkmhlennimv 3060

RESULT 6
Y77905
ID Y77905 standard; Protein: 3060 AA.
XX
AC Y77905;
XX
Db 13-JUN-2000 (first entry)
XX
DE Plasmodium var-7 polypeptide.
XX
KW DBL gene; Duffy-binding like gene; ebl-1; Duffy Antigen Binding Protein;
KW DABP; Stialic Acid Binding Protein; SAbP; malaria; vaccine; immunisation;
KW protozoacide; var-7.
XX
OS Plasmodium sp.
XX
PN US5993827-A.
XX
PD 30-NOV-1999.
XX
PF 07-JUN-1995; 95US-0487826.
XX
PR 10-SEP-1993; 93US-0119677.
XX
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Sim KL, Chitnis C, Peterson DS, Su X, Wellms TE, Miller LH.
XX

```

DR WPI: 2000-194198/17.
 DR N-PSDB: 298287.
 XX
 PT Isolated protein binding domains from Plasmodium vivax and Plasmodium
 PT falciparum erythrocyte binding proteins useful for vaccinating against
 PT malaria -
 XX
 PS Disclosure: Columns 109-124; 93pp; English.
 XX
 CC The invention relates to ebl-1 polypeptides that are encoded by the DBL
 CC (buffy-binding like) gene family. The ebl-1 proteins are substantially
 CC identical to the Duffy Antigen Binding Protein (DABP) and Sialic Acid
 CC Binding Protein (SABP), which are soluble proteins that appear in the
 CC culture supernatant after erythrocytes infected with malaria release
 CC merozoites. Immunohemical studies indicate that DABP and SABP are the
 CC respective ligands for Plasmodium vivax and Plasmodium falciparum Duffy
 CC and sialic acid receptors on erythrocytes. The ebl-1 polypeptides may be
 CC used to vaccinate against malaria, especially caused by P. falciparum.
 CC Immunization with the polypeptide provides effective protection against
 CC malaria. The present sequence represents the var-7 polypeptide.
 CC
 XX
 SO Sequence 3060 AA:
 Query Match 39.7%; Score 4802; DB 21; Length 3060;
 Best Local Similarity 36.2%; Pred. No. 3.1e-281;
 Matches 1156; Conservative 277; Mismatches 631; Indels 1128; Gaps 83;
 QY 4 SGGSGGODEAKVHDEFGQVHDEVHGEAKNVSELKSLASIIIGETAFVTKSQT 63
 DB 12 aagddedesaakimfdrgkdydkveakegkqglqrls-----eakfeknesdp 65
 QY 64 ES-----KYTELIEANSKRNPCK-----KDGKGNVDNRSVYEQAGNKKMKC 107
 DB 66 qtepdpedldtkyhtntctn-vinpcadrsdvrfsdegygqctnrlkdsqgdakg--- 121
 QY 108 SNGMTCAFFRRLALCNKFPNMNSNDSSKAKHDLAEVCMARYEGESIKTHRYKDYSKY 167
 DB 122 ----acapyrrllhvcddqleglqleplkltnt-hnllvdvcmaakefegsltdgpykqytl 176
 QY 168 PGSPFPMCTMLARFADIGDIIRGRDLYLGKNNKKQNGKEFEREKLEOKLKEITRKIIDN 227
 DB 177 gdspsqgclctmlarsfadigdlvirgrdylgnpgeik-----qrqqlennlltflfgklyek 231
 QY 228 LKDRKAKRYNGDEDPNPFYKREDMWTANRETVMGAMTCSKELDNSSFRATCNDTGOCP 287
 DB 232 lngaea--ryg--ndpefiklredwcanetvkalctnaw--gnlyflactn---rg- 281
 QY 288 SQTNNKCRCDKGMAGKPKAGDGVITVPTFVDYVPOYLRFEEWAEDEFCRRKKKLE 347
 DB 282 ertkygcrnddq-----vptyfdyvpqylrfeewwedecrkhkklk 325
 QY 348 NLEKOCGKXKSDERYCSRNGYDCBOTISRKGVKMGKCTDCFPACGSENMIDNOK 407
 DB 326 dvknccrgkykdeddyrcsngydcetkralgklyrgkqisclycnpydvdlmqke 385
 QY 408 QPDRQ-KKYTEKEI-----SDGGRRKKRAGV--TTKREGEKSFYEKLKNDGVTDAF 458
 DB 386 qfdqgkkydeeklykengaaagstrqkrdagylcttnydyekkyfdeinkseyrtvdkt 445
 QY 459 LGLLNNEKACKDITD--GKINFEVNSGGVGGSGTSGASGTNDENKGFYRSEYC 516
 DB 446 leklsneelctkvkdeegtlfdkfvn-----sdstgagsgtveegstfyrekyk 496
 QY 517 QPCDDCGVQ--HKGC--NQMRKTKVKKMKRSKLYK---INGKMLLLKSLAKVVKDMMI 569
 DB 497 qpcpycgkvkvnnggsaneweeekn-gkcskgllyepkpkdegtltilklsqkphdd--- 552
 QY 570 LKKMKKEFCLTNSDSDSVGSVTTGASGNSKKELLYDEKCKYKHNHYOKVNNQGEVEE 629
 DB 553 leeklnkfcdckngdltinsgsgtctgsggnsqtdelyeekcykgeevlvvghdedee 612
 QY 630 DDELKAGAGGLCILPNPKKKEVSEAKSONNHADIOKTFHDFYVVAHMLKDSIHMRK 689

DB 613 dyenwkaagylcilkngkknkeegntsekepedeqtkrfnffyywahnmlkdslnhk-k 671
 QY 690 RLKSCISNGKTKMK-RNCCNNKKCDCEKWKVQKQETEMKPRIDHNRKTOGREGYFTTLE 748
 DB 672 klqrcldgnnrlkcgnnkncndceefkrlwtqkdwgklyqhktlqnlkrgsgsdhae 731
 QY 749 LI-----LKIQFLKED-----TEENTENSUDAEEAEELKHLQKILKLENNLAV 793
 DB 732 lipdhdvlyqnlqgeelkqdsedaseksensldaeaeelkhlrelisedsnngas 791
 QY 794 VNAG-TEQKTLMDKLNHLELNDATKCKDCPLPEEDK----- 828
 DB 792 vggvteqknmdkllyekdeadclleheeeekkgdgnecieegenfrynpsge 851
 QY 829 ----- 828
 DB 852 sgnkrypylankvaygmhhakltqlasragrsalrgdsiaqfkngrngslkqjckln 911
 QY 829 -----SKRGADPS-----PDIFIP----- 843
 DB 912 enyndsrgnsgpctygdghgvyrmrgtewsnieggkqtsyknvllpprrhmetcn 971
 QY 844 ----- 843
 DB 972 lenldvgsvtknkashelldgvglaektlaaealkrykdqgnlqltdpjqkdqeamcr 1031
 QY 844 ----- 843
 DB 1032 avrfsadlgdlirgrdmwdeksldmetrlitvfnlkexhdgikdnpytgdskkp 1091
 QY 844 ----- 843
 DB 1092 aykklradwcanrhqvwramkatkglicpmpvddyipgrlwmtemaewyakaqsge 1151
 QY 844 ----- 843
 DB 1152 ydklkkicadcmksygdgctgtdvdcgkcaacdkyeeleakwneqwrklsdkynllylq 1211
 QY 844 ----- 843
 DB 1212 aktstnprtlvgddpdygmwdfllrlnkasiaarvlvkraagsptelaaaapltry 1271
 QY 844 ----- 843
 DB 1272 staagylngelgygqcgqctfcekknghatststtkenkeytfkqppeyatacdlnrs 1331
 QY 844 ---RPEEKED----- 850
 DB 1332 qteepkkeeenvesaaklvekilegkngrttvgecnpkesyprdkcnldishdgamp 1391
 QY 851 -----DEN-----EDDEDEVRD--- 863
 DB 1392 prqklcilylahesqenlktddnlkdaflktaaeelflsqwykakndseaklldrgl 1451
 QY 864 -----DETRAKE-----TTESG- 875
 DB 1452 lpsqflsmytftgdyrdicntdlskqndvakakdklgykffiskdskspsglsrgew 1511
 QY 876 -----ATDT----- 879
 DB 1512 ktngpelwkgmlcaltkyvtldtnkrklkndysdkvnsgnqgnpsleefaakpqflrmw 1571
 QY 880 ----- 879
 DB 1572 lewgeefoaeqrkhenllkdaacnelnstqgcndakhrcngacraygeyenkkkefsqgt 1631
 QY 880 ----- 879
 DB 1632 nntvikanvpqdpeykygyekdgvqplqngneyllqkcdmksccmgnvlsvapkexf 1691
 QY 880 -----TTSLDVCPIVGKVLTKDNESLQDA 903

D 1692 gkYahkypckcdyqgkhvpsipppppvqpqpeapltvclvdcslv-klflkdtmfda 1750
Q 904 CSLXGNGNNSRLGMKCV---PSSGEPT---SSDKNGALCVPRRRRLYIKIIVMATPT 957
D 1751 c9lky-gktapsawcclpsdtksgagatkyksgsdsgscicprrrrllyvgkligetata1 1809
Q 958 ESPQASGEASSTSTPPDSKFLKAFVESAIETFFLHRRYKEEKAAVOEGAGH 1017
D 1810 --pgsegapnsira-----ddlnafisaatetflwdrkyeeekpgg-dgsqga 1858
Q 1018 LPRVEE--GSPEDYEDK-LKEGKIPDGLRQMFYTLGDRDILFSGSNDTTSVKDRPS 1074
D 1859 lsgltstysddeedppkllqgkllppdflrlmfyltgydrcllvhgn--tsdsngntg 1916
Q 1075 SSNDLKNIVLLASGSTEOERKMKYKEI--KNFRKSTESAPNLVSHPOTWENNGK 1132
D 1917 snnn--llvlaasgnkdmqkigekllgllpknngtclvpyks--aqtpkwwnehæ 1970
Q 1133 YTMGMVCALT---SKOKIAKVEKKPKQIENPENLME-----ANKK-----PKR 1175
D 1971 slwkmicallytcknptdsagde---nkjckdveyekffgscadkngtastptgtyk 2027
Q 1176 POYQTYNKLDENSGTSPTTQTOASDNTPTTLTHFYKRPYTFWFEEMGESFCRERKK 1235
D 2028 tgydyekvklledtsg-----aktpsasadtcllstdfvlrpyfryleewgqnfckkrth 2081
Q 1236 RLKQIRKVDCKYBNDVG-----RCSGDEACDSISTHDYSTVPSENCPCGKHCSSY 1287
D 2082 klagikheckveengsgrgtlrgysgdgaacnemlpkndgtvplekpcakpcssy 2141
Q 1288 RKMIRKKIEFHKOSNAYGGOOKTDATRNNGNPFDEKFCCTLETTWDAKFLRLKNPKK 1347
D 2142 rkwlaskykefekakayeqk-dkcvngsmkhngfceltscakadflctl--gpk 2198
Q 1348 TNKEYGDDIDFEKDSKTFQHTGYCGPCPKFTNCQNGCGVSGLNGNCDGKSIDAKET 1407
D 2199 pnvvegkcltfd--ddkcltkhckdpcplkfsvnckkdec--nsqytdcrnkslatacl 2254
Q 1408 AKMRSTTDVWVRVSDNNTNTEGODLKDACQHANIFKIRBDWKCVCYGCVIDCEQTN 1467
D 2255 engvdstv-lcmrvsadsqfngdglenacrgagllfegllrdewkcnrcvgyvvcckpen 2313
Q 1468 INERTDKEYIOIRALFRKRVNENFLDYKINDKISHCKEGSGKINGCEKNSKCLEK 1527
D 2314 vngeekyghlqiraltvkrveyfedynklkhkshrikngelispcl---kn--cvek 2367
Q 1538 WIEKKIAEMENIKRRFNDOYENKDDPDYNVKSILEBELPKIAVVDODNVKILCVFENS 1587
D 2368 wvdqrrtkewkelterfkdgkndsdndvzsfletllpqlidanakokvklkskfgnsc 2427
Q 1588 GCTLISNTQ--NNKENDAIIDCKLKLGVAKKCPGK--PSEKOSDCKEPPPLPDE-- 1639
D 2428 gcsaanegnkngeykdaicmlklkldigecckkhqtsdctsdclpqtlldecltd 2487
Q 1640 ---EBOGNEENTLEBPKECPPT--TQPREKGEFGKGNKEEKKDEKKESEEPKAEESR 1694
D 2488 ddtleeaakkmm--pklicenvlkaqgedeg--c-----vpaenseepaatdsqk 2535
Q 1695 AAEBPAPTAESEETETNPEPPGTGPAAPPTPAP---PTPDTPLRLRQADPEPDTIL 1751
D 2536 etpeqtlvlpkeeeavppeppp-----ppqekapaplpqppqpqpqpqlldnphvltal 2589
Q 1752 QT-TTPFGVALAGSIALFLKKKTKASYGNLFQILQIPKSYDIPTLKSSNRIPYVSD 1810
D 2590 vstlawsvglgfatfylvllkktksvgnlfgllqpkdsydlptklsparylpytsq 2649
Q 1811 RYKGRYTIEMGDSDDEKAFMSDPTDVSSSEYEELDINIIVYPGSKYKTLLEVLE 1870
D 2650 kyrgkrylyllegdsqtds-gydcnhsdltsseseyeeemindilyvpskyltllievle 2708
Q 1871 P-----SGNNTTASGKNTSPDTRNDIQNDGIPSSKITDNMWNLKKEFTSN 1916
D 2709 psgnnttasgnnttasgnnttasgkntpsdclqndgipsskltldnewnqlkdeffsq 2768

Q 1917 MLONQNDVNDYTSGNSTNTNTTTSRHNVDNNTNTMSRDNMEENLLPSIHGONLY 1976
D 2769 yljsepnlepn-----mlgyndvndmhlctshhveek-pflmslhnrltl 2813
Q 1977 SCEEYSYNV-----NMVN-----SMNDIPINDNNVYSIGDILINDLSGSKPI 2019
D 2814 sgeeynydmfngsnnpindlsdtsnmadsltsnmhpnkndlysgldlindalsgnh-1 2872
Q 2020 DIYDEVLAKRENELFETE-NTRKSTQNVAKTNTSDPILHNOLELHKLLDRHRDMCEKK 2078
D 2873 dlydemlkrkenelfgtkhthkthntynvakparddptlenglnlfbkwldtrhrdmeekw 2932
Q 2079 NKEDIILNKAEMKENINNSGRTYNSDKRPSHNHVLNTDVSIOQDMNPKTRNETTND 2138
D 2933 nherlpklkelw--enechsgdl--nsglpsgnhvlntdvsigldmknpktknelmtmd 2988
Q 2139 TNQDSTMDTIIDLEKYNDPYDFEDDIYHDVDYESSMDIYV'DHNNVTSNNMDV 2198
D 2989 trpdkstmdtliddlekynepyydfyddllyhdvdeksmdlly'dhnnvtnnmdiv 3048
Q 2199 PTKMHIENINV 2210
D 3049 ptkmhlmemlrv 3060
RESULT 7
W00384
ID W00384 standard; Protein; 2913 AA.
XX AC
XX W00384;
DT 21-FEB-1997 (first entry)
XX DE
XX Plasmodium falciparum erythrocyte membrane protein.
XX DE
XX Plasmodium falciparum; erythrocyte membrane protein; malaria;
KW detection; identification; treatment; prevention; parasite.
XX OS
XX Plasmodium falciparum MC type.
XX FH
FH Key
FH Domain
FT 62..394
FT /label= Duffy binding ligand domain 1
FT Region
FT 607..648
FT /note= "Cysteine rich motif"
FT Domain
FT 839..1272
FT /label= Duffy binding ligand domain 2
FT Region
FT 1482..1527
FT /note= "Cysteine rich motif"
FT Domain
FT 1706..2005
FT /label= Duffy binding ligand 3
FT Domain
FT 2102..2349
FT /label= Duffy binding ligand 4
FT Region
FT 2354..2398
FT /note= "Cysteine rich motif"
FT Domain
FT 2450..2475
FT /note= "Putative transmembrane domain"
FT FT
XX W0963736-A1.
XX PN
XX 31-OCT-1996.
XX PD
XX 26-APR-1996; 96WO-US05798.
XX PF
XX 27-APR-1995; 95US-0430908.
XX PR
XX (AFY-) AFFYMAX TECHNOLOGIES NV.
XX PA
XX Baruch DI, Howard RJ, Pasloske BL;
XX PI
XX WPI; 1996-497376/49.
XX DR
XX N-PSDB; T41852.

Dh	1593	cgqnhnpgnktacdpfrspipreestpdladdtprdpdladdqhtceqpfccppppppmtcvek	1562
Qy	1569	-----	1568
Dh	1653	akeIveaegkInmelkngnkdFngcnvkkknngavigeescfegfyensvnmInmc	1712
Qy	1569	-----	1568
Dh	1713	kdnqnerfkIqgkwfkYygtlrkdclirprehmcldlsmIgrtlisdssaIlkIqde	1772
Qy	1569	AVVNDQDNVtKLVFENS-----KGC-----	1589
Dh	1773	aakerdIdlrkllegnsdcdehIcdamkysfadiIdIgrdrIwnkskqglqkrlEY	1832
Qy	1590	-----TLISNTQNNKEND-----	1602
Dh	1833	aflInyIklqndknkYekdrpkyIqIdrsdwGanrkhIwamtcnapddakfIkknpndt	1892
Qy	1603	-----	1602
Dh	1893	sgsssskYImtchncsgYdkpeppdydYlpqpfIrwmqewsesfckllneemeqfektcgec	1952
Qy	1603	-----AIDOMLKKIKGVKAKNCPG-----	1620
Dh	1953	knsItceddrIngtnceencckngceYkklImhwIgfdkYkeIyelnYnmkdsKinsney	2012
Qy	1621	-KPSGERQSD-CRE-----P-----P-----P	1635
Dh	2013	fkkfIeklkdkckelnsdckIdeathcckYkfnsenkhnhnyafnmpkYekackcd	2072
Qy	1636	LPDEEDONPEE-----NTLEPPKFC-----	1657
Dh	2073	apdpldncpksaIyekacntIprtkIceskfInddsdwtsfYqtsprdnTgvlvpr	2132
Qy	1658	-----	1657
Dh	2133	rrqclKnIttklrIslekIddfIaeImtsaYneqkllelykKddvtIqamkysfYdYg	2192
Qy	1658	-----TPOPEERKG-----ETCG	1671
Dh	2193	dIvYgdIdIstaprdklctkclnvIlkYgdgtneIkedrgkwreentItrwamIcgykYkaag	2252
Qy	1672	NKEEKKD-----EKKESEEPKAEESG-----PA	1695
Dh	2253	gkIeerdcslpddntbqfIrwfrewsehIcarkrqkIfevkrccasaqclIeygtIdprv	2312
Qy	1696	AEEBAP-----TAESE-----TETNFPPEPGCGRAP-----	1723
Dh	2313	ceecacYgrdyItrkIqeyrIlInygnvtrInfekaevtIkapEYfIndkndcnclskYId	2372
Qy	1724	-----PSTP-----APPNDPppIPLRQ-----	1741
Dh	2373	IekkwkmYsIddndIknkcIcrtqkIprkrrppkYvYpreeehbpsqdcpprIppkppdl	2432
Qy	1742	--ADEPDSITLOTTIPFGVALAGSIAPFLUKKRTKASVGNLFOIIOIPKSDYDIPTLK	1799
Dh	2433	ppaeepfncrdIektIpfIgaIaIgsIafIflkktksvgnIlgIhIhpsdydIprkI	2492
Qy	1800	SSNVIrYvSDRKkKRTYIYMEGDSDEDKYAFKMSDTDVT--SSEBEYEELDINDIYVGS	1858
Dh	2493	spnYrIytsckYrgkYrIyIYegdsqIdS--gYtIdhysdItsseseYeelIndIYVpS	2551
Qy	1859	PKYTLLEVLIEPsgNNtTASgKNPESDPRNDIONdGIPSSKITTENNOCLKKEPISNML	1918
Dh	2552	pkytllIevlIepsgnmttsqYknpstcdqndIgdngIpskItIcneuntlkdefIsml	2611
Qy	1919	QNOBNdVNDYTSgNSSTNTNTITTSRHANvNNtTMSRONMEENLLPSIHdGNLYSG	1978
Dh	2612	qneuptePn-----mIygvndntIhptshvveekpImsIhdrdIysg	2656
Qy	1979	EEYSYNNWNNNSMNDPIFINdNNVYSGIdLINDLSGKGPIdIYDEVYKKRENEfEGTEN	2038
Dh	2657	eeyYvnnvnnv--ndpIIsarngYsgIdIndIsInsk--vdiYgeIlIkrkenelfgfgh	2713

Oy	2039	TKR-TSQONVAKTNSDPIHNOLELFPKWLDRRRQCEKM--KKNEEDILNKLKEPMNKE	2094
Dd	2714	tktntstnsvaknntclptlmglnlflhtwldlrdrmcckwdcnkkeeildklkeewmkd	2773
Oy	2095	-----NNNSGCTVYSDDKPSPS-----HN---HYLANTDVSIOIDMDNPKRKNIEIT	21359
Dd	2774	nmsgnlmpsgntcrlptpsdlspgksqsdslpsdmnlhsdrlpyllntcvsqtlmndprkpfnfs	28333
Oy	2136	NMDDTNDQKSTWDTLLDDLLEK-YNDPYYYDFEDEDIIYHDVDEKSSMEDIYVDH--NNVT	21922
Dd	2834	nmdcypnssmdtllledikpfnepyyydvqdq--lygvhv-----dhatsvvd	28806
Oy	2193	SNNMDVPFKAHIEKNIVYNKKELFEFEEYPISDITMTI	2228
Dd	2881	tnandepskvyqlend-vnlk--lvkekypdiadlwdi	2913
 RESULT 8 R70236 ID R70236 standard; Protein; 2703 AA. XX AC R70236; XX DT 22-SEP-1995 (first entry) XX DE P. falciparum Proj3. XX KW Erythrocyte binding ligand; Proj3; binding domain; malaria; therapy; KW vaccine. OS Plasmodium falciparum. PN MO9507353-A. PD 16-MAR-1995. PE 07-SEP-1994; 94WO-US10230. PR 10-SEP-1993; 93US-O119677. PA (USSH) US DEPT HEALTH & HUMAN SERVICES. PI Chitluis C, Miller LH, Peterson DS, Sim KL, Su X. PI Wellens TE; XX WPI: 1995-123427/16. DR N-PSDB; Q83529. XX PT New erythrocyte binding domain polypeptide(s) - isolated from PT Plasmodium binding proteins, used in diagnosis, treatment and PT prevention of malaria XX Disclosure; Page 61-65; 81pp; English. XX PS Erythrocyte binding ligand (EBL) family genes were cloned from CC P. falciparum chromosome 7 subsegment libraries constructed from CC genetic studies of the chloroquine resistance locus. The 4 genes, CC EBL-e1 (Q83526), E3la (Q83527), EBL-e2 (Q83528) and Proj3 (Q83529), CC encode the proteins given in R70233-36, respectively. The binding CC domains of such proteins can be expressed e.g. in B. coli, yeast, CC mammalian, insect, and in vaccinia virus and adenovirus-infected CC cells, and provide protection against P. falciparum. XX Sequence 2703 AA;			

	Query Match	26.4%	Score 3197.5	DB 16	Length 2703
	Best Local Similarity	30.6%	Pred. No. 2.8e-184		
	Matches 834	Conservative 244	Mismatches 569	Indels 1075	Gaps 74
QY	4	SGGSGGTODEDAKHYLDEGGQKVVHDEVCEANNVYSELKGSLSLAITLCEETAFYVAKSMOT	63		

Db 7 aagddiledesakmfdtigrkdydkvkeekergkylgrlrs-----eakfeknesdp 60
QY 64 ES-----KYTELEBANSKRNPCK-----KDGKGNVDRFSYKEQAGYDNKKKCC 107
Db 61 qtpepdolchkyhtnvtln-vlnpcadrsdvrfdsyggcethrldksqgdhkyg--- 116
QY 108 SNGMTCAPFRLHLCKNKNFPMNNSNDSSKAKHDLAEVCMAYKESIKTHYPRKYSKY 167
Db 117 -----acapyrrllhvcdqleqlpklfnt-hnllvdvcmakfeggsrltgdpykyaty 171
QY 168 PGSDPMTMLABSPADIGDIRGRDLYLGKNNKKONKKEFERELEBEKLEIKKINDN 227
Db 172 gdspsqctcmrlarfdldgrdlylgnpqlk-----grqlemlnlltlygklyek 226
QY 228 LKDEAOKRYNGDEDPNFYKLREDMWTANREYVNCAMTCSKELDSSYFRATCNDTGCP 287
Db 227 lngaea--ryg--ndpefklrtdewtanretvkwalcna--gnlyfhtcn---rg- 276
QY 288 SQTNNKCRDKDKGAMNACKPRAGDGVITVPTFYDYVPOYLRFWEEMADECRKKKKLE 347
Db 277 ertkyccrncddq-----vptyfdyvpqylrfeeweaderckrkklk 320
QY 348 NLEKQCRKDKSDERYCSRNGYDCEOTISRKKYVRMKGCTDCEFFACGSENMIDNRK 407
Db 321 dvkncrcykdckeddycsrnydccklraigrlykqisclyacnpyvdlmgke 380
QY 408 QFDKO-KKYTKEI-----SDGGRKKRAVGG--TTKEYEGESFYELKNDGYGVDAF 458
Db 381 qfdqgkkydeeklkyengasgsgrqktdagttcttngdyekfyedlnkseyrtvxf 440
QY 459 LGLINNEKACKDITD--GKINFRVNSGGVGGSGTSGASTDENKGTFRYRSEYC 516
Db 441 leklsneelctkvkdeegtlldfkvn-----sdstsgasgtlvegsqgtfyskyc 491
QY 517 QPCDGCQV--HKGG--NOWERKTVKKMRWSKLYKR---INGKVVLLKSLKVVKDMIT 569
Db 492 qpcygcqklyvnnsgsneweekn-gkcskqklyerpxdkqegtlitllkqgqhd--- 547
QY 570 LKRWKRECLTQNSSDGSVVTTGASGNSSEKELDEKCYKHNENVOKVNOGVEEE 629
Db 548 leeklnfcdckngdtlntsgsgtsgsgsgnsgfqlyeekcykgevlvnyvngdedee 607
QY 630 DDELKGAAGLCILPNPKKKEVSEAKSONNHADIQKTEHDFYUWVAMMLKDSIMWTK 689
Db 608 dyenvknaagjcllknqkknkeegntsekerdelqtfnfgyuwahmkdslnhk-k 666
QY 690 RLKSCISDGKMKC--RNCNKKKCDCEKVMVKOKETEMJRINDHKTQGTSEGYUFTTLE 748
Db 667 klqpcldgngrlkcgnkcnndcecfkrltqkxdewgklyqhhtkltqnlkqrggsdntae 726
QY 749 LI-----LKLOFLKED--TEENTENSLDAEAEELKHLKLENNENLAV 793
Db 727 llyrdhdyvlyqnlqgeelkqdsedaeeeksenldeeeaelhlhrelisesdnngas 786
QY 794 VNAG--TEOKTLMKDLKLNHELNDATKCKDCPLREEDK----- 828
Db 787 vvgvteqknimdkllnyekdeadclletheeekkekggnecieegnfryncsge 846
QY 829 ----- 828
Db 847 sgnkrypvlankvaygmhvaklqjastragsalrgdlsiaqfknrgnstlkgjckln 906
QY 829 -----SKGRSADPS-----PDIRP----- 843
Db 907 enynderagngspctygdghgvymrlgtevenslegkkytsknyvllprtrehmcstn 966
QY 844 ----- 843
Db 967 lenldvsvtkndkashnllgydvqjaaktaaeelkrykdqnnlqltdpigrkgqeamcr 1026
QY 844 ----- 843
Db 1027 avrysfadlqdllygrdmwdeksstcmelrlltvcfnlkekhdglykdnpytgdaskkp 1086

QY 844 ----- 843
Db 1087 aykklradwmeanrhqvrnankcatkglicpmpvddyipqrlrwmteawaewyckaqsge 1146
QY 844 ----- 843
Db 1147 ydklklcadomskysgdyctgtgdvdcgckaackkyeeleekwneqwrklisdynllylq 1206
QY 844 ----- 843
Db 1207 aktstspgrtvlgddpdyqmwdfllpjlhkaiaarvlvakraagsptelaaaapltry 1266
QY 844 ----- 843
Db 1267 staagylnqelgygqceqtfceekkhgatssttkenkeytfkqppeyatadclnrs 1336
QY 844 ---RPEEKED----- 850
Db 1327 qteepkckkeenvesacklvekllegkngrtlvgecnprkesydwcdknnldishdagcmp 1386
QY 851 -----DEN-----EDDEDEVRO--- 863
Db 1387 prqklyllylahesqlelnlktddnlkdaflktaaeelflswgyksakndseaklldrgl 1446
QY 864 -----DETAKE-----TTEGS----- 875
Db 1447 lpsqflrsmmytfqdyrdiclnltstskqndvakakdklykfiskdyskspsglsrgew 1506
QY 876 -----ADTF----- 879
Db 1507 ktngpelwgmllcaltkyvtldtnkrlkndysdkvnsqngnpsleefaakpqflrwm 1566
QY 880 ----- 879
Db 1567 lewgeefaerqkkenllkdaacnelnstqgcndakhrcnagrayqeyvenkkkefsqgt 1626
QY 880 ----- 879
Db 1627 nmlfkanvqdpdeykyeykdygvnqigneyllqkcdmkscomgnvlsvapkexf 1686
QY 880 -----TSLDVCPIVGKYLTKDNESLADA 903
Db 1687 gkyahkyepkcdcygklyhvpelppppvpqpqpearlvtvdcslv-klfkddmnsda 1745
QY 904 CSLKYCGNNSRLGRCV---TPSGEPTT---SSDKNGACVPPRRRLYIKIYDMATKT 957
Db 1746 cglky-gktapswkclypsdtksgagatgksgdsqslcipprrrrllyvqkigewatal 1804
QY 958 ESPQSGSEASTSGSTTPPDSEALKAFAVESAAIETFLMHRKBEKKAVAOEGAGHG 1017
Db 1805 --pqgegaaphrsra-----ddlrfnafgsaatefflwdrykeekkpqg-dgsqga 1853
QY 1018 LPRVBE--GSPETYPEDK-LKEGKIPDGFLKQMYTTLGDIYDLIFSSNDOTTSVKOTPS 1074
Db 1854 lsqllstysddeepbpkllyngkllppdflimlytlylgydydllyvhgn--tsdsnglmg 1911
QY 1075 SSNDNLKNIVLASSTGEQREKKNKYKEI--KNFRKSTFRSAPNLVSHQOTWENNGK 1132
Db 1912 snm---nlvleasgnedmqklegklqglpknggrlpykps---aqpcdkwneae 1965
QY 1133 YIWHGVCALF---SKDKIAKGYEKKPOKLENBENLME---ANK-----PKP 1175
Db 1966 slwgmicalyteknpdtsargde---nkekdeveyekfsgstakghglastrpityk 2022
QY 1176 POYOTYNNKLDENSGTSBRTTQOASSDNTPTTLTHFVKKPPTYRRMTEBEGESCRERKK 1235
Db 2023 tgydyekvklledtsq-----aktpsaasdlp-llsdfvlrppryfryleevgnqclckrh 2076
QY 1236 RLKQIKYDKCYENDDVG-----RCSGGEACDLSISTHDYSVRPFNCPGCGKHGSSY 1287
Db 2077 klqikneckveenggsstrggltrqysgdgeacnemlpkndgfvpllekpscakpssy 2136

Qy	1288	RKMIERKKIEFHQSNANAYGGQKTDARNRNNGNFPDKEFCFQTTLETWDAKFLERLKNPGCK	1347
Db	2137	IKwIesKgefeKqekayeqqk-dKcvnsnkhhdngfcelIttsakdfIKLl--gpck	2193
Qy	1348	TNKEYGDDIDIFEDKSTQFQTEYECPCPKFNTNCGNGCGSGVANGNCDDGKSIDAEKI	1407
Db	2194	pnavegItfd---dktfItnktdccpCIkIsynckkdecd-nskgtdcrnknsIdatdI	2249
Qy	1408	AKMRSSTTDVVMRVSDNDNTFEBDDLKDACOHANIFKGIKRDVVKCGYVGVDCICEQTN	1467
Db	2250	engvdstv-IemrvsadsksgefnngdgIenaocrgagIfegIrtdewkcrnvegYvvcKpen	2308
Qy	1468	INENTDKREYIQTIRALFKRVNVEFLBDYKINDKTSIHCLKKEGSKCINGCEKSKCLEK	1527
Db	2309	vngaeakKhIhIqIraIvkrvveyfIdynkIkKhIshrlkngelIsPcI---kn--cvek	2362
Qy	1528	WIEKKIEMENIKRRFNDQYENKDDPVNWKSTILEELIPKIAVVDODNVIRIKLGFENSK	1587
Db	2353	vwdkIrKwKeteIerfIdqyKndnsddadvstfletIlIpqIdanakKhvIkIskfgnsc	2422
Qy	1588	GCTLTISNTQ--NNKENDAIIDCMKLKLGVRAKNCPGK--PSCGKOSDCKEPPPLPDE---	1639
Db	2423	gcsasanegqkngelydaIdcmkIkIkdIgecekKhbqtstcecdIprgqtIedeld	2482
Qy	1640	---EDQNPENTLEPPKFCPPT--TQPPEKGGETCGNKEEKDKKESEBPAAKESGP	1694
Db	2483	ddIteeaKkhnm--pKlCenvIkItagqdegg--c-----vpaenseepatdsqk	2530
Qy	1695	AAEPAPATASEETENNFPETGPAAPPSIPAP---PTPTPTPLPLPQADEPDSITL	1751
Db	2531	etpqrIvIkpeeeavpeppP-----ppqekapapIpqppIcpIqIIdnpvItal	2584
Qy	1752	QT-TIPGVALALGSIAFLFK	1772
Db	2585	vstIawsvglgfatfIyfIyIk	2606
RESULT	9		
W22482			
ID	W22482	standard; Protein: 2710 AA.	
XX			
AC	W22482:		
XX			
DT	07-OCT-1997 (first entry)		
XX			
DE	Plasmodium Proj3.		
XX			
KW	DBL gene family; SGBP, stalic acid binding protein; vaccine; therapy;		
KW	Duffy binding like gene; Duffy antigen binding protein; erythrocyte;		
KW	DAMP; merozoite; malaria; var-1; var-2; var-3; var-7; immune response;		
KW	Plasmodium.		
XX			
OS	Plasmodium falciparum.		
XX			
PN	MO9640766-A2.		
XX			
PD	19-DEC-1996.		
XX			
PE	07-JUN-1996; 96MO-US09508.		
XX			
PR	07-JUN-1995; 95US-0487826.		
XX			
PA	(USSH) US DEPT HEALTH & HUMAN SERVICES.		
XX			
PI	Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;		
XX	Wellens TE;		
XX			
DR	WPI: 1997-052231/05.		
XX			
XX	N-PSDB; T72897.		
PT	New malaria vaccines - contains cysteine-rich DBL family protein		
	binding domains homologous domains of the Duffy and stalic acid		
	binding proteins		

[illegible]

QY	630	DDDELKAGGGLCILLPNPKKNKEVSEAKSQNNHADIOKTRHDFPFYVWAHMLKDSIHMRWK	689
DB	615	dyenwknagglcillknqkknkeegntsekepelelqctfnpfyvvahmlkdsihw-k	673
QY	690	RLKSGISGKTKMKC-RNGCNKKCDCCFFKWKQKQETEMKPIKOHFKQOEIPRGYFTTLE	748
DB	674	kIqfclqgnrikcgmnkcmndcecfIrwIqtqkdgwgkIvqhfktqnlkgr:sgsdntae	733
QY	749	LI-----KLOFLKED-----TEENTENSLAEAEELKHLOKLTENNNLAV	793
DB	734	llpfchdyvlyqnlgeefllqgdsedaaseeksensldeaeeekhlreliesednnqas	793
QY	794	VNAG-TEOKTLMIDLNLHNELDATKCDCPLBEDK-----	828
DB	794	lvvgvteqklnmdkllnyekdeadlclclhedeeeekedydgnecleegentfrypcsg	853
QY	829	-----	828
DB	854	sgnkrypvlankvaygmhkhaktqlaasragralrgdlsaqfkngrngstlkqjckln	913
QY	829	-----SKRGRAADS-----PDRFIP-----	843
DB	914	enysndsrngsgprcctqkdqdhgvgzmrljgtewsnlegkktcsyknvliprrhmtcsn	973
QY	844	-----	843
DB	974	lenldvsgvtckndkashnlllgdvqlaaktdaaelkrykdqmnlgltqrlqgkdgemcr	1033
QY	844	-----	843
DB	1034	avrysfadlgdlirgrdmwdecksstmetrllltvfknkkehndgllnphklytgdeskkp	1093
QY	844	-----	843
DB	1094	aykkIradwneanrhqyvrzmkcaatkylilcpgmpvddyIprlrlmwteaweyckasqse	1153
QY	844	-----	843
DB	1154	ydkllkIcadcmksqgdgkctqgdvdcgkckaacdakykeelekwneqwrIsdkynllYlq	1213
QY	844	-----	843
DB	1214	aktetpgrtvljgddpdyqgmwdfllplkhaslaarylvkrraagspteIaaapltcy	1273
QY	844	-----	843
DB	1274	staagyIhgeIgygscqeqtqfcekkhgatsstctkenkeylftkqpppeyatacdlnrs	1333
QY	844	--RPEKED--	850
DB	1334	qteepkkeeenvesaackIveklllegkngrtltvgcenpkasesydwcdckmldshdgacmp	1393
QY	851	-----DEN-----EDDEDEVD--	863
DB	1394	prrqkIcIyIahesqtenlktdldnlkdaeflswgyykskndseaklldrgl	1453
QY	864	-----DEETAKE-----TEGS-----	875
DB	1454	lpsqflrsmnytfgrdyrdiclntcdIskknqndavakakdgIgfIfskdsgkspsglstrgsw	1513
QY	876	-----ATDT-----	879
DB	1514	ktngpelkqgmIcaltktyvcdtnkrrklrkdysdydkvngsqngnpsleefaaqpflrwm	1573
QY	880	-----	879
DB	1574	IewgeefcaerqkkenllkacneInstqgcndahnrncqacraygeyenkkkefsgqt	1633
QY	880	-----	879
DB	1634	mnfvlkanvqpdpbeykyeyekdgvpqrlqgneylIqkcdnkcscmdgnlvsvspkepf	1693

QY	880	-----TSLDVCPLVGVNLRKDNESLQDA	903
Db	1694	gkyahkypekcdcygkghvpsipppppyvqgpearpvavdcslv-kltfldtnfnfsda	1752
QY	904	CSLKYGGNNRSLRGVCV---TPSGEPT---SSDKNGAICVPPRRRRRIYIKIYDMATKT	957
Db	1753	cgiky-qltapsawkcipdsctsgagatcgsgdssticlprrrrllyvgklqdegatal	1811
QY	958	ESPQASGEASSTIGSTTTPPDOSKDALIKAFVESAAIETFTLWHRKREKKAVAGGAGHG	1017
Db	1812	--pqgeagaphsra-----ddlrnatlfgaaletffldxykeekbpqg-dgsqga	1860
QY	1018	LPRAVE--GSEYVPEEK-LKEGKIPDGFRLQMFYVLGRDILSGSNDTYSVKQPS	1074
Db	1861	lsgtscydeedepbkllqngklppdflrmlfyldgrydllyhngn--tsdsgntng	1918
QY	1075	SSNDLKNVLLASGSTEOEREKMNRYKEI--KNFRKSTERSADNLVSHQTMWENNGK	1132
Db	1919	snmn---nlvlaasgnkedmqklqekleqllpknngtclpivpss---aqtpdkwnehae	1972
QY	1133	YIWHGMNCAFL---SKDKTAKGVKEKKPKQIENPELME-----ANKK-----PKP	1175
Db	1973	slwgmcalalyteknptdsargde---nlekddeevyekffgstadkngtasrptglyk	2029
QY	1176	POYOYTNKLDENSGTSPRTTQOASSDNTPTTLTFHFVRPTYFPMFEMESFCREKK	1235
Db	2030	tgdyekwkleatdsq-----aktpasasthp-llsdfvlrppfyfryllewgnfcfckrkhh	2083
QY	1236	RLKQIKKDKCKYENDVG-----KCSDDGACBDSISHDYSTVPSRNCGCCGHCSSY	1287
Db	2084	klagklhckcyeeenggsrrrgtilytsgydgaaenemlpknqgtyvpdlekpscapecssy	2143
QY	1288	RKWLERKIEFHKOSNAYGOOKTDATRNNGNTFDEKFCFTLTETWPAKFLERLKNQPC	1347
Db	2144	rkwlcsygeteketkeqakeyegq-dkcvnsnkhndgfcetltssakfdiktl--gpc	2200
QY	1348	TNKEYGGDDIDFEKDSFTFOHTEYCGPCPKFTKNCQNGCGVSGLNGMCDGKSIDAKEI	1407
Db	2201	pnveygkllfcd---ddkrlfthktcdcpclkfsvnckkdecd--nskyltclcrknsidatcl	2256
QY	1408	AKMNSTTDVMRYRSDMTWTFEGDDUKADCOHANFKGIRKDVKCGVCGVDICEQTN	1467
Db	2257	engyvdstv-lemrvsadsksqfngdglmaocrgaqifleglrxidekkcnvcegyvavckpen	2315
QY	1468	INERTDGKEYIQIRALPKRWENFLBEDYNNKINDKISHCIKKGESKCINGCEKNSCKLEK	1527
Db	2316	vngaeqkghlqldalrylkyrweyffedqylkhlkshlrxkngelspcI---kn--cvek	2369
QY	1528	WIEKKIAWEMIKRRFNDQYENKDDQDYNNKSLTELTPKIAVWVDQDNVTKLCVFEKSK	1587
Db	2370	vwdqrfkewkeilerfkdygkhdnsdddvrslelllpqldanakhvklakfngsc	2429
QY	1588	GCTLTISMTQ--NNKENAIDCMKLKGVAKNCPCK---PSGEKOSDCKEPPPLPDE---	1639
Db	2430	gcsaanegknngykaldaicmklkldkltgecekkhhqtsctescdpqpqtiledeld	2489
QY	1640	---EDQNPENTLEPPKFCPPPT---TOPPEBKGGETCGNKEEKDKKESSEPAKESGP	1694
Db	2490	ddlsteaekkhmm--pklcenavlktaqgedeg--c-----vpaensepaatdsqk	2537
QY	1695	AAEPAPPAESEETETNFPPPGCTGAAPSPAP---PTPTPPPLPQADDEPDSITL	1751
Db	2538	etpctqvyllkpeeaavpeppp-----ppqekapaplpqpqptcptqclldnphvltal	2591
QY	1752	QT-TIPFGVALALGSIATFLFK	1772
Db	2592	vtslawsvgigfatfeyfjlk	2613
RESULT	10		
ID	Y77904	standard; Protein: 2710 AA.	
XX			

AC Y77904;
XX 13-JUN-2000 (first entry)
DE P. falciparum Proj3 binding domain polypeptide.
XX DBL gene; Duffy-binding like gene; ebl-1; Duffy Antigen Binding Protein;
KM DABP; Sialic Acid Binding Protein; SABP; malaria; vaccine; immunisation;
KW protozoacide; Proj3.
XX Plasmodium falciparum.
OS US5993827-A.
XX 30-NOV-1999.
XX 07-JUN-1995; 95US-0487826.
XX 10-SEP-1993; 93US-0119677.
XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
PA Slim KL, Chitnis C, Peterson DS, Su X, Wellens TE, Miller LH;
PI WPI: 2000-194198/17.
DR N-PSDB: 298266.
XX Isolated protein binding domains from Plasmodium vivax and Plasmodium
PT falciparum erythrocyte binding proteins useful for vaccinating against
PT malaria -
PS Disclosure: Columns 79-92; 93pp; English.
XX The invention relates to ebl-1 polypeptides that are encoded by the DBL
CC (Duffy-binding like) gene family. The ebl-1 proteins are substantially
CC identical to the Duffy Antigen Binding Protein (DABP) and Sialic Acid
CC Binding Protein (SABP), which are soluble proteins that appear in the
CC culture supernatant after erythrocytes infected with malaria release
CC merozoites. Immunochemical studies indicate that DABP and SABP are the
CC respective ligands for Plasmodium vivax and Plasmodium falciparum Duffy
CC and sialic acid receptors on erythrocytes. The ebl-1 polypeptides may be
CC used to vaccinate against malaria, especially caused by P. falciparum.
CC Immunization with the polypeptide provides effective protection against
CC malaria. The present sequence represents the Proj3 binding domain
CC polypeptide.
XX Sequence 2710 AA;
SO

Query Match 26.4%; Score 3197.5; DB 21; Length 2710;
Best Local Similarity 30.6%; Pred. No. 2.8e-184;
Matches 834; Conservative 244; Mismatches 569; Indels 1075; Gaps 74;

QY 4 SGGSGAGTODEDEKAVLDEFGQKVEHGEAKNYVSLKSGSLASILGETAFVYSMOT 63
DB 14 aagggddiedesakhmfdrigkdvkveakeergkylgrls-----eafeknesp 67
QY 64 ES-----KYTELLEANSKRNPCK-----KDGKNDVDRFSVKEQAGYDNKKMK 107
DB 68 qfpepdcdldhkyhvnvtn-vlnpcadrsdvfsdeygsgcthmrlkdsqgdnkq--- 123
QY 108 SNGMTCAPPRRHLCKNKNPNMNSNDSSKAKHDLAEVMAKYEBSKITHPKYDSKY 167
DB 124 ---acapyrrlhvcdqnlqepiklnt-hnllvdvmaakfegqsitgdpykqaly 178
QY 168 PSDEPMCTMLARSPADIGDIRGDLVGNKKKONGKETREKLEOKLEIFKKIHDN 227
DB 179 gdspsqicmlarstfadvgdlygrdlygnpeik-----grqlennlklftgkiyp 233
QY 228 LKDKERQKRYNGDEDPNFYKLRREDWMTANRETFVGMATCSKELDSSYFRATCNDTGOCP 287
DB 234 lngaea--ryg--ndpeffklrldewltanrelvkwaitcnaw--gntlyhacn---rg- 283

QY 288 SOTHNKRCDKDKGANAKPKRAGDDVITVPTTFDVPOLRMFEEMAEIDCFKRRKKKLE 347
DB 284 ertkyrcrondq-----vptyfdvpylrvfwewaedfcrkknkik 327
QY 348 NLEKOCRGKDKSDEYRYCSRNGYDCEQOTISRKGVRMGKCTDCEFFACGSYENWIDNQR 407
DB 328 dkrcnrcgdkedkdrycsrnygdcektkralgklyrgqcslyacnpydwlnnqke 387
QY 408 QPDKQ-KKTYKEI-----SDGGRRKRAVVG--TTKBYGKRSYELKKNYGTVDVAF 458
DB 388 qtdkqkkkydeekikyengasgsrqkdaagttlttnydyekkyfde.lnkseyrtvdx 447
QY 459 LGLNNKRCADITD--GGKINFEVNSGGGVVGGSGGTSCASGNIENKCTFPRSEIC 516
DB 448 leklsneictkvkdeegttidfkvn-----sdsstgsqgn.esggtfyrskyc 498
QY 517 QPCPDGVQ--HKGG--NOMERKTRVKKRWSKLYP---INGKVLJLKSLLVYKDMVI 569
DB 499 qpccpygyvkvnngsgsnweekn-gkcksgklyepkpdkegttltlksqgknhd--- 554
QY 570 LKKNKRECFLTNNSDGSVGSVYTTGASGSEKKELYDEMKCYKINEYQKXNVGGEVDE 629
DB 555 leeklnkfcdenkngdtlinsgsgtigsqgnsgqrqelyeewkcyxgedvkvghdedee 614
QY 630 DDELKAGAGLICILPNPKKKEVSEAKSONNHADIQKPHDFEYVWVAHMLKDSIHMRK 689
DB 615 dyenvknagglcilkngkknkeegntsekepedeqktlnpflywamldsdshmk-k 673
QY 690 RLKSCISDQKTWKC--RNGCNKKDCFEKWKOKETEMKPIKHFYTCBGIPIGYYFTTLE 748
DB 674 klgrelqgnrlkcgknknndcecfkrlwtgkdwgmlyvghfktqlnkygsgndtae 733
QY 749 LI-----LKLQFLKED-----TEBNTENSIDAEAEELKHLKIKLENNENLAV 793
DB 734 lpfndhdyvlynlqeeftkgsdesaesensidaeeaelkhl::eliesednqeas 793
QY 794 VNAG--TEOKTLMDKILNHELDATFKCQDPLPEDEK----- 828
DB 794 vvggytlegknmdkllynkedeadicliehedeekekydgnecl segenfrynpcsg 853
QY 829 ----- 828
DB 854 sgnkryvlaankvaygmhbkaktqlasragrsalrgdsiaqfngngstlkgqlckin 913
QY 829 -----SKGRSADPS-----PLIFTP----- 843
DB 914 enysndrsngsgpctgkdgdhgyvmrlygtewsnlegkktqsykrvflprprehmcen 973
QY 844 ----- 843
DB 974 lenldvsgvtnkdashsllgvdqlaaktaaelkrykdnqlqldp!gqkdqeamcr 1033
QY 844 ----- 843
DB 1034 avrrysfadtgdllrgrdmwdeksstmetrlltvtfknlkehgdg!kdnpytgeskkr 1093
QY 844 ----- 843
DB 1094 aykklradwneanrhqvramkcatkyl!cpmpvddy!pqr!rvntewaewyckasqe 1153
QY 844 ----- 843
DB 1154 ykllklicadsmksqdkctgtdvdcgkkaacdyykeeliekwegwrkisdynll!y!q 1213
QY 844 ----- 843
DB 1214 aktstngtrtv!gdddpdygmwdf!lp!hhaslaar!v!vraasgspteiaaap!p!y 1273
QY 844 ----- 843
DB 1274 staagy!hngel!ygsgceqet!fcekhhgatsctckenkey!ffk!pppey!acac!nrs 1333
QY 844 -----REPERED----- 850

Dd	1334	qteppkktkeenevesacklvekliegkngrttvtegcnpkspcydwckdmkldshdnagamp	1393
OY	851	-----DEN-----EDDDEDEYRD--	863
Dd	1394	prqkklclylahesqtenlktddnlkdaefiktaaaetflswgykkskndseaklldrgl	1453
OY	864	-----DEETAKE-----TMEGS-----	875
Dd	1454	lpsqflsmyntfgydldclndlskknqndvakaakdkgkfksdgakspsgsrgeww	1513
OY	876	-----ATDT-----	879
Dd	1514	ktngpelwkgmllcaltkyvelctdnkrtklkndysydkvnsgnqngpsl eefaaqpqlrwm	1573
OY	880	-----	879
Dd	1574	lewgeefeaerqkkenlllkacnelnstqcgndakhrncnagcraygeyvenkkhefsgqt	1633
OY	880	-----	879
Dd	1634	mfvkanvqpbqdbeykygywkdgvqplgneyllqkcdanlkcsomdgnvlsvapkekf	1693
OY	880	-----TSLDVCPIYGVKYLTKDNESL0DA	903
Dd	1694	gkyahkyepkcdcygqkhvpslpppppyvqpgppearltvtdvcsrlv-klfllfdlnfsda	1752
OY	904	CSLKYGCNNSRLGRCY---TPSGEPYT---SSDKGALICVPRRRRLYIKKIYDMATKT	957
Dd	1753	cgkly-ghtpswskcrlpsdtksagaatgksgsdsgslcldprrrrrlyyvgklgwatal	1811
OY	958	ESPOASGSBASSTSGSTTPPDOSKEALLKAFVESALIEFFLHRRKKEEKKAADOEGAGHG	1017
Dd	1812	--pgagaaapshsra-----ddlrnaflfgsaatetflldxyrkeekbpqg--dgsqga	1860
OY	1018	LPVREE--GSPEYDEPK--LKEGKITPDGFLROHFYILGROYILSGSNDTPYSKDNPS	1074
Dd	1861	lsqtlstysddeepprpkllgnkllppdlrlmtfylvlgdyrtllhyngn--tsdagngn	1918
OY	1075	SSNDLKNIVLLASGSTOEREXMKNYKEI--KNFRKSTERSADPNLVSHPOTWENNKG	1132
Dd	1919	snmh--nlvleasngkmedmqkldqekleqllknpgnrlpvpss--aqtpdkwnehae	1972
OY	1133	YIWHGMVCALT---SKDKTAKGVKKPKQIENPEULME-----ANKK-----PKP	1175
Dd	1973	slwgmicallyteknkpdtsargde--nklexddevyekffgstadknhtasrlptgtyk	2029
OY	1176	POYOYTNVKNLDENSGTSPRTTOTQOASDNDPTTLTHFVPRPYFFMFEWGESFECREKK	1235
Dd	2030	tgdyevkylkledtsq-----aktpssasdtpr-llsdfvlrpyfrylleewgnfcfkxkh	2083
OY	1236	RLKQIKYDKCKYENDVG-----KCSGDGEACDSISTHDIYSTYPSFENCPCGCKHGCSY	1287
Dd	2084	klaqlkheckveeenggsrrrgylltgyrsdgaeacemlprkndgtvpydlekpcakepsasy	2143
OY	1288	RKMJERKKIEFHKOSNMYGOOKTDATRNNGNPTDFKFCFTLETFWMDAKFLPERLKNKGSC	1347
Dd	2144	rktvlesygtelckekakeyegqk-dkcyngsnkhndngfscelttsakadflkltl--gpcok	2200
OY	1348	TNKEYGGDDIDFEKDSKTFQHTTEYCGPCPKFKTNGCNGNGCVSGJNGNCGDKSDAIKEI	1407
Dd	2201	pnavegkrlfifd---ddktcfhktkdcprclksfynckkdecd--nskytdornkhsidatdl	2256
OY	1408	AKMSSSTPDVVMRYSDDMTFEGDDLKAQACQANIFKGIKIRDWKKGVCYGVGDIQEDTNI	1467
Dd	2257	engvdsrlv-lemrysadsaksgfngdglmaecpagaflleglrldevckcnvogyvvyvckpen	2315
OY	1468	INERTDGEKEYLOIALFKRMVNEFLLDYDKINDKISHCIKSGEGRKJCINGCEKSNKCLEK	1527
Dd	2316	vnsgaekghhllqalralyktwveyffedynaklkhhkshtlkngeisprcl---kn--cwek	2369
OY	1528	WIEKKAEMENIKRRFDYENKQDPDYNVKSTLEELLPKIAVVDQDNVTKLCFVENS	1587

Db	2370	wvdkirkwexkterfddqyknhdnsdddnvrstlecllpqldanahkvklkifgnc	2429
Qy	1588	GCTLISTNQ--NNKENDATICMLKRLGVAKKCPGK---PSEKSDCKEPPPLPDE---	1639
Db	2430	gcsasaneqknqngykaaiclmlklklkdkgceekhhqtsdtecsdpqpfledelid	2489
Qy	1640	---EDONPEENTLPPKFCPPPT--TQPPPEKGETCGNKEKKDKKESEEPAAKEESGP	1694
Db	2430	dditeeekkrhmm--pplcenvlvtktgdeegq--c-----vpaenseepaatdsqk	2537
Qy	1695	AAEEDAPTAESEETETNFPPEPGCTGAAPSPAP---PTPTPPPLRQADEPPDSTIL	1751
Db	2538	etpeetpylkrpeeeavdeppepp-----ppqekapaplpqpppctpcqjldnphvltal	2591
Qy	1752	QT-RTPEGVALAGSIAFLFK	1772
Db	2592	vstclawsvglgfatfctfytlk	2613
RESULT	11		
ID	W00385	W00385 standard; Protein; 1726 AA.	
XX	AC	W00385;	
XX	DT	21-FEB-1997 (first entry)	
XX	DE	Truncated Plasmodium falciparum erythrocyte membrane protein.	
XX	XX	Plasmodium falciparum: erythrocyte membrane protein; malaria; detection; identification; treatment; prevention; parasite.	
OS	XX	Plasmodium falciparum MC type.	
XX	FH	Key	Location/Qualifiers
XX	FT	Domain	62..394 /label= Duffey binding ligand domain 1
XX	FT	Region	607..648 /note= "Cysteine rich motif"
XX	FT	Domain	839..1282 /label= Duffey binding ligand domain 2
XX	FT	Region	1488..1523 /note= "Cysteine rich motif"
XX	XX	W09633736-A1.	
XX	PD	31-OCT-1996.	
XX	PF	26-APR-1996; 96MO-US05798.	
XX	XX	27-APR-1995; 95US-0430908.	
XX	PA	(AFFY-) AFFYMAX TECHNOLOGIES NV.	
XX	XX	Baruch DI, Howard RJ, Pasloske BL;	
XX	XX	WPI: 1996-497376/49.	
XX	XX	N-PSDB: T41853.	
XX	XX	New Plasmodium falciparum erythrocyte membrane proteins - used to develop products for the diagnosis, treatment or prevention of malaria parasite infections	
PS	PS	Claim 1; Figure 12; 149pp; English.	
CC	CC	A polypeptide comprising a Plasmodium falciparum (Pf) erythrocyte membrane protein 1 (PfEMP1) or active fragments or analogues of that protein can be used in the treatment or prevention of symptoms of a malaria parasite infection. The polypeptides can inhibit, block or reverse the sequestration of erythrocytes in patients suffering from malaria. Nucleic acids derived from the PfEMP1 gene can be used as probes and primers to identify a Plasmodium falciparum parasite, the primers used to generate characteristic amplification patterns from	

Qy	1120	VSHQTFWNNNGKTYIHGMYCALTSKDTAKVYKKPKIENPMLMBEANKKPKRPOY	1179
Db	988	---qefwgtcygkdiwkgmicalq-----	1020
Qy	1180	YTNKCLBENSQTSRPTQTOASSDNTPTLTHPVKRPYERMFEEGMSFCERRKRLQ	1239
Db	1021	ysnvtftghl-----	1055
Qy	1240	IKVJCKY--ENGDNVRCSGGEACDSISTHDYSTVPFSCPCGHCSYRKMIERKIE	1297
Db	1066	lkercmwygyngdkdgykdkkek-----	1105
Qy	1298	FHKOSNAYGOOK--TDATRNNGNTFDEKFCCKLETWMDAKFLER--LKNGPCKTKERYGSD	1355
Db	1106	ytkqngytlewkygsjpkedsdvckeaky-----	1151
Qy	1356	DIDFKNSKTFQHTVEYCGPPCPKFKTKCQNGCNGVSSGLNCCGDKSIDAKELAKRSSTY	1415
Db	1152	dl-----ayc-----nemeq-----tst	1165
Qy	1416	DYVARVSDNDNTFEGBDLKDACAQAHANIFKGIKRDYWKGCYGVGDICEQTNINERTDGK	1475
Db	1166	d-----sn-----	1169
Qy	1476	EYIOIRALFKRWENFLEDYKINDKISHCIKKGESKICNGEKNKSCLEKWIETKIAE	1535
Db	1170	-----ndnlpselk-----	1178
Qy	1536	WENIKKRFNNOYEEKODPDVNVKSLIEBELLPRKAVVNDQDNVYKLCVENSNGCLISNT	1595
Db	1179	-----yppl-----eleegct-----	1189
Qy	1596	QNNKENDAIQMLKLGKAKNCPGKPSGKSDCKEPPP--LPDEEDONPEENTL--EP	1651
Db	1190	-----ckdpspgveip--ekkppekvipkr	1213
Qy	1652	PKFCPPPTQPPPEKGETCGNKEKKKDEKKESEEPKAEESGPAAEBAPPAESEETETN	1711
Db	1214	pkl--pkrpker-----	1225
Qy	1712	FPEPPGTPGAAPSTPAPRPPTDTPPLRQADPEPSTLQTTIPPGVALAGSIAPLFL	1771
Db	1226	fp-----tpa-----lknmlsctimsigifactfyly	1256
Qy	1772	KKKTKASGNLFOILOPKSDVDIPLTKSSNRIPVSDRYKGYKTYIMEGSDSDDKAF	1831
Db	1257	kkktkst--dllyvlnpkssdydipkllspuryiplysgkygkyrylyegsgtcs--gy	1314
Qy	1832	MSDPTDVT--SESESEBELDINDIYVGSFKYKTLLEVLYLEPSGNWTTASGKTPEDTEND	1890
Db	1315	tdhysdltsseseeysaelndindiyaprapkyklllevlylepsgmnttsgmtpsdgind	1374
Qy	1891	IQNDGIFSSKITDNEWNLKKEFISNMLON--QPNDPNDYTSGNSSTNTNTTTSRHNVD	1949
Db	1375	lqndgipskkitdnewnlkdefisqylqseqndvpndysgsdiprlbtq-----	1424
Qy	1950	NNTWTTMSRDMENLILPSIHGONLYSGEBSYNNWNNNSMNDPIRINDNNVYSIDLI	2009
Db	1425	--pntlyfndpkepfitsiherdlysgseeyynnmvntndnplsgknglysgidll	1481
Qy	2010	NDSLSGKPIDIYDEVLKREKENELFCTENPKR--TSTQNAKATTSNDPIHNOLEPHKYL	2068
Db	1482	ndslsnsm--vdiydevlykrtenelefctlnhtkhtcnsrvakelcgqblmngidllhkwld	1540
Qy	2069	RHRDMCEKMKNEEDILNKLKEENKE-----NINNSKG-----TYN	2104
Db	1541	rhrdmcekmkneevldklkeewnkdmnsgnlnpsgnlnpsgntrcptsdipsgklsdtps	1600
Qy	2105	SDNKRPSNNHVLNLDVSTQIDMDNPKRKNETLNTMDTQDKSSTMDTILDLDEKXNDPYRYDF	2164
Db	1601	dnalpsnktlntlntvstqlmndpkrplngftn-----ndtllledlekxyney--d	1649

```

OY      2165 YEDDIIY--HVDVDEKSSMDDIYDVHNNTSNMNOVPFEMTHLENNIVNKKKEIFEEBEPVI    2222
          ||| | | | | | :                || ||||| : ||| | | | : | : ||| | |
DB      1650 VQDDIIVdhdvhtagta-----snamdvskvgiemdl-nlk-lyxekypil    1694

OY      2223 SDIWNI 2228
          ||| : |
DB      1695 sdvwgi 1700

RESULT  13
R70235
ID      R70235 standard; Protein; 700 AA.
XX
AC      R70235;
XX
DT      22-SEP-1995 (first entry)
XX
DE      P. falciparum EBL-e2.
XX
KW      Erythrocyte binding ligand; EBL-e2; binding domain; malaria; therapy;
KM      vaccine.
XX
OS      Plasmodium falciparum.
XX
PN      WO9507353-A.
XX
PD      16-MAR-1995.
XX
PF      07-SEP-1994;   94WO-US10230.
XX
PR      10-SEP-1993;   93US-0119677.
XX
PA      (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI      Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;
PI      Wellens TE.
XX
DR      WPI: 1995-123427/16.
DR      N-PSDB: Q83528.
XX
PT      New erythrocyte binding domain polypeptide(s) - Isoated from
PT      Plasmodium binding proteins, used in diagnosis, treatment and
PS      prevention of malaria
XX
PS      Disclosure; Page 56-57; 81pp; English.
XX
CC      Erythrocyte binding ligand (EBL) family genes were cloned from
CC      P. falciparum chromosome 7 subsegment libraries constructed during
CC      genetic studies of the chloroquine resistance locus. The 4 genes,
CC      EBL-e1 (Q83526), E3ia (Q83527), EBL-e2 (Q83528) and Proj3 (Q83529),
CC      encode the proteins given in R70233-36, respectively. The binding
CC      domains of such proteins can be expressed e.g. in E. coli, yeast,
CC      mammalian, insect, and in vaccinia virus and adenovirus-infected
CC      cells, and provide protection against P. falciparum.
CC
SQ      Sequence       700 AA;

Query Match           11.4%; Score 1385; DB 16; Length 700;
Best Local Similarity 40.8%; Pred. No. 1.7e-75;
Matches 328; Conservative 106; Mismatches 207; Indels 162; Gaps 34.

OY      113 CAPFRRLHCNKNPMMNSNDSSAKKHDLIAVCMAAKYEGBSIR'HPKYDSKYPGSDP    172
          |||:|||||: | : | : | | | | | | | | | | | | | | | | | : :
DB      10 caprrrlhncdy---nlesidltstckllilevcmaakeynsl'nhyvqhgarrtredssas    66

OY      173 PWCMTAMANSFADIGDIIRGRDYLG--NKKRKQNGKETEREKLEQILKEIFEKIH-DNLK    229
          |||:|||||: |||:|||||: |||: | | : | | | | | | | | | | | | |
DB      67 qlctvlarstafdigtlivrgkdlilydnkexeq-----rkklleg:lkdifkllkdwymk    120

OY      230 DKEAQRINGD-EDDPNFYLREDMWTPANRETYWGAMTCSKELDNS:YFRATCNDTGQGPS    288
          ||| : | : | : ||||| ||||| | | : : : : | | : | |
DB      121 tngaqeryyddakggdfqgliredwvtswnetetwkalichapkean/fikltaen-vykg--    177
```

```
QY 289 QTHNRCRCDKDKGANAGKPKAGDGVITVPTVFDVVPOLYLRMFEEWADFCRKKKKLENN 348
    178 -tngqchc-----1ggd-----vpcyfdyvpqyltwfeewaefctrkkkkkllen 220
QY 349 LEKQCRGKDKSDEYRYCSRNGYDCEOTISRKQVBMKGCTDCFCFAGCSYENWIDNQRKQ 408
    221 lqgqrdyegn---lycsngnydccktllykqkvlvghehcncswcmymetwldnqke 277
QY 409 FDKOK-KYTKELISDGG-----RKKRAVGTTKYE--GYEKSFEYKLLKNDGTVDAFLG 460
    278 flkqkrkyetelsggsgskspkrctraarsasssdngyeskfykklveygqvdcktlk 337
QY 461 LLNNEKACKDITDGGKINFKFEVNSGGVVGSGGSGTSGASGTNDENKGTFFRSEYCCQPCP 520
    338 llnkeglcqkqpyg--nekadn-----vdfnekyvtfstfctelcepcp 380
QY 521 DCGVOHKSGNOMERTKYKKRMWSKLYRPINGKMWLLSKLVKDMMLKKNMKEPCLT 580
    381 wcggle-kggppw--kvkgdktcgsaekctydpknltidprvlypdksgqnlkkyknfc-- 435
QY 581 QNSSDGSVGSVVTTCGASGNSSEKKELYDEMKCYKHNEVQKNNVQGEVEEDDELKAGAGL 640
    436 -----ekyapqpgqqltk-----wcgy-----dehr----- 456
QY 641 CILPNPKKNKEVSEAK-----SONNHADIOKTFHDFEYVVAHMLKDSIHWRTKRLKSCI- 695
    457 ---psaknnnncevgtwckftgskgt--vkaynvfwvdmldhdsvewkte-lskcln 510
QY 696 --SDGKYTKCKRNGCKKDCFEKVVYKOKETEMKPKIDHFKTOEGT-----PEGY 743
    511 nntgnlctornnhckctdcgcfqkvwckqgemwmaikdhfgyqtdlvqgkglvlfsgy-- 568
QY 744 FTYLELIK-----LOFLKEDTEENTENSLDAEAEELKHLKILKENENNLAVNAGTE 799
    569 --vidlvkgnnllqnlk-dvhgdc-----dalkhikl--ldeedavavv199kd 614
QY 800 QKTLMKDLINHELNDATCK---DCPLPEDKSRGSRADSPDIFRPP-----E 846
    615 nttl-dkllqhekegaecqkqkececkkaqgesrgsraetredertqtpadsagevee 673
QY 847 EKEDDENEDDEYRDDEETAK 869
    674 edddydedddvvgdvavse 696
Db
RESULT 14
W22481
ID W22481 standard; Protein; 700 AA.
XX
AC W22481;
XX
DT 07-OCT-1997 (first entry)
XX
DE Plasmodium ebl-2.
XX
KW DBL gene family; SABP; stailic acid binding protein; vaccine; therapy;
    Duffy binding like gene; Duffy antigen binding protein; erythrocyte;
    DABP; merozoite; malaria; var-1; var-2; var-3; var-7; immune response;
    Plasmodium.
XX
OS Plasmodium falciparum.
XX
PN W09640766-A2.
PD 19-DEC-1996.
XX
PE 07-JUN-1996; 96WO-US09508.
XX
PR 07-JUN-1995; 95US-0487826.
XX
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
```

```
PI Chitnais C, Miller LH, Peterson DS, Sim KL, Su X;
PI Wellens TE;
XX
DR WPI: 1997-052231/05.
XX
DR N-PSDB: 1772896.
XX
PT New malaria vaccines - contains cysteine-rich DBL family protein
    binding domains homologous domains of the Duffy and stailic acid
    binding proteins
XX
PS Disclosure: Page 46-48; 96pp; English.
XX
CC This sequence represents ebl-2 of Plasmodium. Ebl-2 belongs to
    the duffy binding like (DBL) family of genes which have homology to the
    duffy antigen binding protein (DABP) and stailic acid binding protein
    (SABP) conserved regions (see 772885 and 772888 respectively). The var
    family of genes modulate cytoadherence and antigenic variation of
    Plasmodium infected erythrocytes. SABP and the duffy antigen binding
    protein (DABP) are soluble proteins that appear in the culture
    supernatant after infected erythrocytes release merozoites. DABP and SABP
    mediate the binding of merozoites and schizonts to the erythrocyte
    surface. These proteins are necessary for erythrocyte invasion by the
    parasite. This sequence can be used in the compositions of the invention.
    The compositions are for the treatment and prevention of malaria, and
    comprise either a nucleotide sequence or encoded polypeptide of the
    var-1, var-2, var-3 or var-7 genes of the DBL gene family, a family of
    genes having homology with conserved regions of DABP and SABP. The
    compositions are used for the treatment and prevention of malaria. They
    are also used in the preparation of vaccines for inducing a protective
    immune response in a mammal to Plasmodium merozoites (especially
    Plasmodium falciparum or Plasmodium vivax).
XX
SQ Sequence 700 AA;
XX
Query Match 11.4%; Score 1385; DB 18; Length 700;
    Best Local Similarity 40.8%; Pred. No. 17e-75;
    Matches 328; Conservative 106; Mismatches 207; Indels 162; Gaps 34;
QY 113 CAPRRLILCKNKPNNMNSKRAKHDLAEVCAAYEEESIKTHYPRKDSYRPGSDF 172
    10 caprrllhcdy---nesldtstclhlllevcaayegsnslctlyqtrtnedaa 66
QY 173 PMSWTIARSPADIGDIINGRDLYLIG--NKKRKQNGKETEREKLDQKLEIRFKTH-DNLK 229
    67 qlcvtlarstadiqdivrgkdllygdnkeqeg-----rkllegklkldfkklhdkwm 120
QY 230 DKEAKRYNGD-EDPNFYKLRDWMWTANRETVWGAAMTCSKELDNSSYFRATCNTDGGSPS 288
    121 tngqeryldakg9dffqiredwtsnretcvkallchapkeanyfllktaen-vgk-- 177
QY 289 QTHNRCRCDKDKGANAGKPKAGDGVITVPTVFDVVPOLYLRMFEEWADFCRKKKKLENN 348
    178 -tngqchc-----1ggd-----vpcyfdyvpqyltwfeewaefctrkkkkkllen 220
QY 349 LEKQCRGKDKSDEYRYCSRNGYDCEOTISRKQVBMKGCTDCFCFAGCSYENWIDNQRKQ 408
    221 lqgqrdyegn---lycsngnydccktllykqkvlvghehcncswcmymetwldnqke 277
QY 409 FDKOK-KYTKELISDGG-----RKKRAVGTTKYE--GYEKSFEYKLLKNDGTVDAFLG 460
    278 flkqkrkyetelsggsgskspkrctraarsasssdngyeskfykklveygqvdcktlk 337
QY 461 LLNNEKACKDITDGGKINFKFEVNSGGVVGSGGSGTSGASGTNDENKGTFFRSEYCCQPCP 520
    338 llnkeglcqkqpyg--nekadn-----vdfnekyvtfstfctelcepcp 380
QY 521 DCGVOHKSGNOMERTKYKKRMWSKLYRPINGKMWLLSKLVKDMMLKKNMKEPCLT 580
    381 wcggle-kggppw--kvkgdktcgsaekctydpknltidprvlypdksgqnlkkyknfc-- 435
QY 581 QNSSDGSVGSVVTTCGASGNSSEKKELYDEMKCYKHNEVQKNNVQGEVEEDDELKAGAGL 640
    436 -----ekyapqpgqqltk-----wcgy-----dehr----- 456
```

```

Db 436 -----ekgap99gq1k-----wgcyy-----dehr----- 456
QY 641 CILPNPKKNKEVSEAK-----SONNHADIOKTRHDPFYVVAHMLKDSIHMRTRKLSCT- 695
Db 457 ---psknnnncevgwtwdfgkgt--vksyvnffwdvwhdmhdsvewkte-lskcin 510
QY 696 --SDGKTMKCRNGCNKKCDCEFEKAWKOKETEMKPIKDHFTQEGI-----PEGY 743
Db 511 mntngntcrnmkcktdcgcfqkvwkqgemmaikdhfgktdlvqkqglvlfspyg-- 568
QY 744 FTTLELIK----LQFLKEDTEENTENSLSAEAEELKHLKOKILKLENNLAVNAGTE 799
Db 569 --vldivlkggnllgnik-dvngdt-----ddikhikkl--ldeedavavvlgkxd 614
QY 800 QKTLMKLLNHLNDATKCK----DCPLPEEDKSRRSADPSPDIFIRPP-----E 846
Db 615 nttl-dkllqhekeaqeqkqkecekkagqesigrsaetredertqpadagsaveee 673
QY 847 EKEDDENEDDEVRDDEETAK 869
Db 674 eddddydeddddvvdvase 696

RESULT 15
ID Y77903 standard; Protein; 700 AA.
XX Y77903;
AC Y77903;
XX 13-JUN-2000 (first entry)
DT
XX
XX P. falciparum ebl-2 polypeptide.
KM DBL gene; Duffy-binding like gene; ebl-1; Duffy Antigen Binding Protein;
KM DABP; Sialic Acid Binding Protein; SAMP; malaria; vaccine; immunisation;
KM protozoacide.
XX
XX Plasmodium falciparum.
XX
XX US5993827-A.
XX
XX 30-NOV-1999.
PD
XX
XX 07-JUN-1995; 95US-0487826.
PE
XX
XX 10-SEP-1993; 93US-0119677.
PR
XX
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PA
XX
XX Sim KL, Chitnis C, Peterson DS, Su X, Wellem's TE, Miller LH;
PI
XX
XX WPI, 2000-194198/17.
DR
XX
XX N-PSDB; 298285.
PT
XX
XX Isolated protein binding domains from Plasmodium vivax and Plasmodium
PT falciparum erythrocyte binding proteins useful for vaccinating against
PT malaria -
XX
XX
XX Disclosure: Columns 67-72; 93pp; English.
XX
XX The invention relates to ebl-1 polypeptides that are encoded by the DBL
XX (Duffy-binding like) gene family. The ebl-1 proteins are substantially
XX identical to the Duffy Antigen Binding Protein (DABP) and Sialic Acid
XX Binding Protein (SABP), which are soluble proteins that appear in the
XX culture supernatant after erythrocytes infected with malaria release
XX merozoites. Immunochemical studies indicate that DABP and SABP are the
XX respective ligands for Plasmodium vivax and Plasmodium falciparum Duffy
XX and sialic acid receptors on erythrocytes. The ebl-1 polypeptides may be
XX used to vaccinate against malaria, especially caused by P. falciparum.
XX Immunization with the polypeptide provides effective protection against
XX malaria. The present sequence represents the ebl-2 polypeptide.
SQ Sequence 700 AA:

```

```

Query Match 11.4%; Score 1385; DB 21; Length 700;
Best Local Similarity 40.8%; Pred. No. 1,7e-75;
Matches 328; Conservative 106; Mismatches 207; Indels 162; Gaps 34;

QY 113 CAPERLHICNNFNPNNDSSSKAKHDLAEQMAKYEESIKRYHOKYDOKYSGSPF 172
Db 10 caperrllhcdy-----nlesidststkhllvecmakyeqnsinthvqndrtedss 66
QY 173 PMCTMLARSPADIGDIIRGRDLYLG--NRKKRONGKETREKLEOKLKEIFKKH-DNLK 229
Db 67 qlctvlarstadiqdlvrgkdllygydnkeq-----rkkleqlkldlffkklhdkvmk 120
QY 230 DKEAQKRYGND-EDPNFYLRDPMWTANRETYWGAMTCSELDNSSYFATCNDGOGGS 288
Db 121 tngaqerylddakgydflqrlredwtsnretwkalichapkeanyfllktaen-ygkg-- 177
QY 289 QTHNKCRCDDKDGANGAKRKAGDGVITVPTFYDVPYQYLRMPFEWAHDFCRRKKKKLEN 348
Db 178 -tngqchc-----lsgd-----vpcyfdvppqylrvfeewaeftcrkkkkllen 220
QY 349 LEKOCRGKDKSDERYRCSNGYDCEQOTISRKGVRMGKCTDCEFCACGYENWMDNQRO 408
Db 221 lqkqcrdyeqn---lycsngydcctlykqklvighctncswcmymetcdnqkke 277
QY 409 FDKOK-KYTKETISDGGG-----RKRAGGTTKYE--GTEKSFYEKLNDGIGYDAFLG 460
Db 278 flkqkrkyetelsigsgsgskprtkraarsssdngyesfkykllhevgyqdvdkflk 337
QY 461 LNNKAKACDITDGGKINKEVNSGGVGGSGGTSAGSAGTNDKMKTFYSEYCOOPC 520
Db 338 llnkegicqkqpvq--nkdadn-----vdfnekyvlftstlecepp 380
QY 521 DCGVQHGKGNQWERKRYKVMKSKLYKPIKNGMVLKSLKVVKDMHILKRNKKEFCLT 580
Db 381 wogle-ksgppw--kvkgkctgsaktkydpnltdipvljpdksqpnllkkykncf-- 435
QY 581 QNSDGSVGSVVTGASGNSSEKKELYDEMCKYKHNQVKNVQGVJEDDELKAGAGCL 640
Db 436 -----ekgap99gq1k-----wgcyy-----dehr----- 456
QY 641 CILPNPKKNKEVSEAK-----SONNHADIOKTRHDPFYVVAHMLKDSIHMRTRKLSCT- 695
Db 457 ---psknnnncevgwtwdfgkgt--vksyvnffwdvwhdmhdsvewkte-lskcin 510
QY 696 --SDGKTMKCRNGCNKKCDCEFEKAWKOKETEMKPIKDHFTQEGI-----PEGY 743
Db 511 mntngntcrnmkcktdcgcfqkvwkqgemmaikdhfgktdlvqkqglvlfspyg-- 568
QY 744 FTTLELIK----LQFLKEDTEENTENSLSAEAEELKHLKOKILKLENNLAVNAGTE 799
Db 569 --vldivlkggnllgnik-dvngdt-----ddikhikkl--ldeedavavvlgkxd 614
QY 800 QKTLMKLLNHLNDATKCK----DCPLPEEDKSRRSADPSPDIFIRPP-----E 846
Db 615 nttl-dkllqhekeaqeqkqkecekkagqesigrsaetredertqpadagsaveee 673
QY 847 EKEDDENEDDEVRDDEETAK 869
Db 674 eddddydeddddvvdvase 696

```

Search completed: May 1, 2001, 13:07:07
Job time: 300 sec

• • •

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 1, 2001, 13:02:07 ; Search time 23.39 Seconds
(without alignments)
1829.918 Million cell updates/sec

Title: US-09-508-967-1

Perfect score: 12100
Sequence: 1 MATSGSGGTQDDAKHVLD.....YNNKKIEFEPEYPSIDYINI 2228

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 185757 seqs, 19210857 residues

Total number of hits satisfying chosen parameters: 185757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents, AA:*

- 1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
- 2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
- 3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
- 4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
- 5: /cgn2_6/ptodata/2/1aa/CTUS.COMB.pep:*
- 6: /cgn2_6/ptodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4817	39.8	2182	2	US-08-487-826B-16
2	4802	39.7	3060	2	US-08-487-826B-14
3	3197.5	26.4	2710	2	US-08-568-459A-12
4	3197.5	26.4	2710	2	US-08-487-826B-12
5	1385	11.4	700	2	US-08-568-459A-10
6	1385	11.4	700	2	US-08-487-826B-10
7	702	5.8	921	2	US-08-568-459A-8
8	702	5.8	921	2	US-08-487-826B-8
9	418	3.5	1435	2	US-08-568-459A-4
10	418	3.5	1435	2	US-08-487-826B-4
11	379.5	3.1	362	2	US-08-568-459A-18
12	379.5	3.1	362	2	US-08-487-826B-30
13	360.5	3.0	411	2	US-08-568-459A-19
14	360.5	3.0	411	2	US-08-487-826B-31
15	345.5	2.9	749	2	US-08-568-459A-6
16	345.5	2.9	749	2	US-08-487-826B-6
17	315	2.6	1115	2	US-08-568-459A-2
18	315	2.6	1115	2	US-08-487-826B-2
19	315	2.6	1115	6	5198347-6
20	311.5	2.6	1663	5	PCT-US93-07261-16
21	288	2.4	1588	5	PCT-US93-07261-11
22	262.5	2.2	3969	4	US-08-061-376-5
23	226.5	1.9	1183	2	US-08-447-031A-2
24	220	1.8	2843	1	US-07-741-940-2
25	220	1.8	2843	1	US-08-289-548A-2
26	220	1.8	2843	1	US-08-452-654-2
27	220	1.8	2843	2	US-08-370-235A-2

28	217	1.8	1706	2	US-08-459-568-2	Sequence 2, Appl
29	217	1.8	1706	2	US-08-399-411-2	Sequence 2, Appl
30	217	1.8	1706	3	US-08-516-859A-2	Sequence 2, Appl
31	217	1.8	2842	1	US-07-741-940-7	Sequence 7, Appl
32	217	1.8	2842	1	US-08-289-548A-7	Sequence 7, Appl
33	217	1.8	2842	1	US-08-452-654-7	Sequence 7, Appl
34	217	1.8	2843	1	US-08-452-655B-2	Sequence 2, Appl
35	217	1.8	2843	1	US-08-452-655B-7	Sequence 7, Appl
36	217	1.8	2843	4	US-08-450-582-2	Sequence 2, Appl
37	217	1.8	2843	4	US-08-450-582-7	Sequence 7, Appl
38	217	1.8	2973	2	US-08-821-355A-7	Sequence 7, Appl
39	217	1.8	2973	2	US-09-003-687A-7	Sequence 7, Appl
40	217	1.8	2973	4	US-09-136-605-7	Sequence 7, Appl
41	216.5	1.8	1719	2	US-08-459-568-4	Sequence 4, Appl
42	216.5	1.8	1719	2	US-08-399-411-4	Sequence 4, Appl
43	216.5	1.8	1719	3	US-08-516-859A-4	Sequence 4, Appl
44	213.5	1.8	2353	4	US-09-377-155-33	Sequence 33, Appl
45	213.5	1.8	2353	4	US-08-913-942-4	Sequence 4, Appl

ALIGNMENTS

```

RESULT 1
US-08-487-826B-16
; Sequence 16, Application US/08487826B
; Patent No. 5993827
;
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chlntis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhuan
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
;
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobb Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487, 826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2182 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
;
; US-08-487-826B-16

```

Query Match	39.8%	Score 4817	DB 2	Length 2182
Best Local Similarity	45.9%	Pred. No. 0		
Matches 1096	Conservative 298	Mismatches 607	Indels 388	Gaps 93
QY	4	SGGSG-----GTOP-----	EDAKHYLDERGQVHNE--VHGAKNMYSLKSGSLAS-I	50
DB	11	SGGSSGKAKKKTSETIYVSDAKDLDRYGEYVEEYKAYAGAKKYTEALKGNLTANR	70	
QY	51	LGEFTAVTSMOTESKYTELLEANSKRNCKKDGKGNDRPS-----	-YKEQNG	99
DB	71	SEETASSIETCTLVKEYERYVNGDGRHRCRDKANEDVNRSDTLGGCTYNRIRKDSQ	130	
QY	100	YDNKKKCSNGMTCAPFRRLHLCKNFPNMNSNDSSAKAHDLAELYCAAKYEGSIKTH	159	
DB	131	GNKV-----GACAPYRLHLCDY--NLSEIDTSTHKLLECYMAKKEGSSINTH	181	
QY	160	YPKYDSKYGSPGPFQCMYLAHSFADIGDIRDRDYLIG--NKKKONGEYREKLEOKL	217	
DB	182	YVQHORTNEDSASQCTYLAHSFADIGDIRDRDYLIG--NKKKONGEYREKLEOKL	235	
QY	218	KEIFKKIH-DNKKDKKXAKRYNGD--EDPFFYLRREDMWTANRETYWAMTCSKEKLDNSY	275	
DB	236	KDIFKKIHDVAKTNGAOGRIYIDAKGGPFLQREDMWTNSNETYWKALICHAPREANF	295	
QY	276	FRATCDTGGGSGQTHNKCRCOKDKGANAAGKFRAGDGYTVPTFYDYPOYLPMPEMA	335	
DB	296	ITYAC-VGK-----TNGQCHC-----IGG-----VPTFYDYPOYLPMPEMA	335	
QY	336	EDPCRRKKKKLENLEKOGCKDKSDERYCSRNGYDCBOTISRKKRVBGRKCTCQCFPC	395	
DB	336	EDPCRRKKKKLENLEKOGCKDYBON--LYCSNGIDCTKTYIRKGLVYGBHCTGCSWC	392	
QY	396	GSYENWIDNQKQFDFOK-KYTKETISDGG-----RKRKRVGTTYE--GYEKSFEYKL	447	
DB	393	RAYETWIDNQKKEFLKQKRYETELISGGSGSPRTRKAAANSSSDNGYESFYKYL	452	
QY	448	KNDGVTUAFGLILNNEKACDITDGKINREYVNSGGGVGGSGGTSGASGTNDENK	507	
DB	453	KEVGQVDYDFLKLINKEICOKOPVYG--NKKADN-----VDFFNEKYV	495	
QY	508	GTFYSEYQOPDPCGVQHKGNOMERKRYKVKMRMSKLYKPINCKMYLLKSLVWDM	567	
DB	496	KTFSEICEPCPCGLE-KGGPFW-KVKGDKTGSAAKTYIDPKNITDIPVLVLPDSQ	552	
QY	568	MLTKNMKEFCULTONSSDGSVGSVYTTGASGENSEKELYDMWKYCKHNEVOKVAVGCV	627	
DB	553	QNLKKKYNFC-----EKAGPGGQJLK-----WQCY--	580	
QY	628	EDDDELKAGAGLCLIPNFKKKEVSEAK---SONNHADLOKTHDFEYVVAHMLDS	683	
DB	581	-----DEHR-----PSSKNNNCCVEGTWDFTOGQOT--VKSNVFPMVDVHMLDS	626	
QY	684	IHMRRKRLKSCI---SDGKTMKCRNGCNKKKDCCFEKVYQKQKTEMKPIKDKHKYBGT--	738	
DB	627	YEMKTE-LSKCTINNNTNGATCRNNKCKTDCGCFKRWWEKKQOEWMAIKDHHGKOTDIYQ	685	
QY	739	-----PEGYFTTLELILK-----LOFLKDETEENTENSJDAEEAEELKLOKILKE	786	
DB	686	OKGLIVFSYSG---VLDLYLKGMLNLQNIK-DVINGDT-----DDIKIKKL--LD	729	
QY	787	NENNLAVNAGTEQKTLMDKLLNHELINDATYCK---DCPLFEEDKSIRGSADSPDIFI	842	
DB	730	EDDAVAVVLGKDNNTTI-DKLQHEKEQAEQCKQOECEKKAQOESRORSARETREDERT	788	
QY	843	PPP-----EKKEDDENEDDDEVARDBDETKAKETTESGATDTTISLIV-----	885	
DB	789	QOPADAGEVEEEDDDDDVEDDEDVVQEEBEKE--EGTVEYTEVTEVEETVEQ	846	
QY	886	-----CPITYKVLTDNLSLADQSLKY--GGNSRLQMRVCTPGSEPTSSDKKALCY	938	
DB	847	EGVKRCODIVGK-LFEDDKSLKACGLKIGPGSKKEFPNNKCYTPSGVSTATSGKGAICV	905	
QY	939	PPRRRLVIKKIVDMATKTESPQASGSEASSTSGSTTPPDK--EALLKAFVESAAIET	995	

Db	906	PRRRRLVGLSQWASH-----GGETTEVSEATSAPOSESEKLTATIESALLET	959
Qy	996	FELMIRYEEKKAAVA--OBGAGHGLPRVEEGSPEDDEKLKE-GKIPDGLQMEVYTLGD	1053
Db	960	FELMHKRYEEKKPRATODGAGLVSLEPSPREDEPQTOLOQGVIPROFLQOMYTLAD	1019
Qy	1054	YKDILFSGSNDT--SVSKDTPSSNDNLKNIIVLLASGSTEODEKRN---KYKEIKNR	1108
Db	1020	YKDILYSGSNDTSPDTTKQTPSSNDNLKNIIVLEASGSTEODEKMKQIOAKIKLNGA	1079
Qy	1109	KCSFRSAPNLVSNP--QTMWENNCKYIWMHOMCALSKOKIAKGVK--PKQIENPENLM	1166
Db	1080	TSGVPRVTKNSVKRPPQOTWMENELAKDITWAMVALYTKENDAGTSAKLEOKKDKLKLWM	1139
Qy	1167	DEANKPKRPPOYQYTNKLDENSGTSPRTQTOASSDNT--PTTLTHFKRPTPYRMFEE	1224
Db	1140	DEANKNTPEIKYQTYTNKLEDES-----AKSNDTIQPTTLKNEFIEITPFRMLHE	1190
Qy	1225	WGESFCRERKRLKQIYKDVCKVENGDVRCSGDGEACDSISTHDYSTVPSFNCPCGKHC	1284
Db	1191	WMSNCFERARRLKQIKHECDEDEGE--KQYSGGCEELFSGYQVNLQDLS--SSCACP	1248
Qy	1285	SSYKWTIERKRIIEHKOSNAYGOQKT-----ATRNGTPEKCEKTKLEWPPAA	1333
Db	1249	RLYKWTIEKKTEYEEKOAKYEOOKSVYENEOKKOCOTOSNNNA--NEFSRLGASPPAA	1306
Qy	1336	KFLLELKGPKTKMEY--GGDD--IDEPKDSKFTFOHTEYCGPCPKPKTKCQNGNCGVGL	1392
Db	1307	EFLQKL--GSKKNDNGYENGEDNKKIDPKNDKTFKHAHSCDPCPTGVACQNGHC--VQSA	1362
Qy	1393	NG-NCDGKSIDAKEIAKMSSTTDVVMVRSNDNTWFE--GDULKDACOHANIEFKIRK	1449
Db	1364	NGEKKCNK--TTAEDIKNKTDPMCNIMVYSDSTWTFELGD----CKSSGIEFKIRK	1417
Qy	1450	DVMKCGYVCGDTC--EQTINENETDGEKEIOTRALFKWENLEFEDYKINDKISHCIK	1507
Db	1418	DEMKCANVCGVDITTEKKIKINGOEGKRYTKELKMLLEFLEEDYNRIKKIKLICTK	1477
Qy	1508	KGEGSKINGCEKNSKCLEKMKIEKILAEWENIKRRNDQYENKDQYVWKSILEELPK	1567
Db	1478	KEDGCKCIKG-----CIERKVOEKTEKMOKINDTYLEOYKND-----GATLNFLEQ	1522
Qy	1568	IAYVNDODNTIKLC---VFENSKGCTLISNTONKKNENDALCMKIKLIGVAKKNCGRPS	1623
Db	1526	FQYRTFENKNAIKPCDGLDQFKTSGLSNSTONGNNDVLCILNKLOKISECKEHS	1585
Qy	1624	GKQSDC-----KEP---PRLDEBDONPEBNTLEPKFCPEPTTQPREBEKGE--TCG	1671
Db	1586	GOTQPCPNSSLSGKEESTLVEDVDYDEEONP--BNKYEOPRCFCDMKEPKKENDDEVGTCG	1644
Qy	1672	NKEEKDKEKESEEPAKESGPAEERAP--FAESEBETN--PREPCTGAAPSPRA	1728
Db	1645	GDEERK--KVEDSVIEQKEEBASAPESPPLTPREPKKEENVPKP-----	1690
Qy	1729	PTPTPTPP-----LPPQADEP--FDSTIQTTIPGVALALGSIATLELTKKTKASVGN	1781
Db	1691	-----PPKKRRKRTKTNVLDHRAVIRPALMSSTIMMSIGIGFAAFYFYLKTKKTSVGN	1744
Qy	1782	LFQILQIPKSDYDIPFLKSSNRRIPIVSDRYKGYITYWEGSD--EDKAYFSDTDTVS	1840
Db	1745	LFQILQIPKSDYDIPFLKSSNRRIPIVSDRHKKGYITYWEGSDSGGEKAYFAFMSDITDTS	1804
Qy	1841	SESEYEELDINDIYVPSPKYKTLIEVLELPSGNSNNTAGSKMTPSDTRNDIOND---GIP	1897
Db	1805	SESEYEELDINDIYVPSPKYKTLIEVLEPS-----KRDTONDHNDIPSDIP	1853
Qy	1898	SS-----KITDENMQLKEFISNMLOPNADYNDYTSGNSSTNNITTTSHRYDNTNT	1953
Db	1854	NSDTPPTTIDDENMQLKDDISNMLONTONTPEH-----ILHDVNDNTH	1898
Qy	1954	TTMSRDNNEENMLPSITHDNLSGGEYSYNV-----NNVN-----SMNDIPI	1966

Db 1899 PTMSRHNDQKPFMSIHDRNLFSGEERYNYDMFNSGNPDISDSTNSMSTLNNHSPY 1958
QY 1997 NRDNVYSIGDILINDSGCKPIDIYDEVLRKRENELEFGETMTKRTSTON--VAKTNSD 2054
Db 1959 NKNNDYSIGDILINDALSGNH--IDYDEMLKRENELEFGETQHPKRTITSNRYVOTQTSDD 2017
QY 2055 PIHNOLELFHKMLDRHRODCERKMKNEKEDILNKLEEMKNEKINSKGTYSNPKPSHNV 2114
Db 2018 PTTNOINLFHKMLDRHRODCERKMKNEHRLPKLELM--ENETHSGDT--NSGIPSGNHV 2073
QY 2115 LMTDVSIGIDMNPRTKNETMDTNOQSTMDTILDLLEKYNDDPYDYEDDIIYHDV 2174
Db 2074 LMTDVSIGIDMNPRTKNETMDTNOQSTMDTILDLLEKYNDDPYDYEDDIIYHDV 2132
QY 2175 DYKSSMDIYDHNNTYSNNDVPTKMHIEKNYNNKKEIEEPEPIS 2223
Db 2133 NDKKASEDHINDHNMKNNSDPTNOVIEKNVINN--OELLONEYPIS 2180

RESULT 2
US-08-487-826B-14
; Sequence 14, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sam, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelien, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3060 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-487-826B-14

Query Match 39.7%; Score 4802; DB 2; Length 3060;
Best Local Similarity 36.2%; Pred. No. 0;
Matches 1156; Conservative 277; Mismatches 631; Indels 1128; Gaps 83;

QY 4 SGGSGTODEDAKHYLDFGQGVHDEVEHGEAKNYVELKSGSLASLIGETAFTVSMQT 63
Db 12 AAGGDIDEDSAKHMFDRIQKDYDVKVKEAKERKGLGRSL-----EAFKEKESNP 65

QY 64 ES-----KYTELIEANSKRNPCK-----KDKGNDVDFRSFBOAGYDNKMKK 107
Db 66 QTPBEDCDLDHKYHTVNTN--VINPCADRSYVRSDEYGCOTLHNRITKSOGDNKG--- 121
QY 108 SNGMTCAPRRLHLCKNKPNNNSDSKAKHDLAELVMAKREGEKTKTHYPKYDSKY 167
Db 122 ----ACAPYRLHYCDQNLQEQIEPIKITNT--HNLVDVCMAMKFBEGSITQOYPKQATY 176
QY 168 PGSDPFCMTMLARSPADIDDIIRGDLVLYGNKKKKONGKETEREKLEK/LKLEIFKKIHON 227
Db 177 GDSPOICTMLARSPADIDDIYGRDLVLYGNPOEIK-----OROOLEINMLKTFKRIYER 231
QY 228 LDKKEQKRYNDEDEDFKFLREDDWTARERVWGMGMSKELDINS--FRACNDGQOP 287
Db 232 LGAEL--RYG--NDEFFKLEDDWTARERVWKAITGNAM--GNTIFHATCN--RG- 281
QY 288 SOTHNRCQDKDKGANAGPKAGDGVTVPTFYFDVVPQYLWFEEMABDFCRKKKKKLE 347
Db 282 ERTKGYCRNDQ-----VPTFYFDVVPQYLWFEEMABDFCRKKKKKLE 325
QY 348 NLEKQCRGKDSDEYRYCSRNQYDCBOTISRKGVBMGCTDCFPFAJGSEYENMDNQK 407
Db 326 DYKRNCRGKDEKDRYCSRNQYDCBOTISRKGVBMGCTDCFPFAJGSEYENMDNQK 385
QY 408 QFDKQ-KKTYTEL-----SDGGKRAVVG--TTKYGEYKSFYEKLKNDGYTVADF 458
Db 386 QFDKQKRYDEELKYYENGASGSRQKRDAGTNTNYGEEKFTELKSEYRVDFK 445
QY 459 LGLNNEKACKDITD--GGKINPKEVNSGGGVGSGSGTSGASGTPDNKGTFYSEVC 516
Db 446 LEKLSNEELICTYVKDEGGTIDFKNPN-----SDTSGASGTPVSGTFTYKSKC 496
QY 517 QPCPDGVQ--HKGG--NOMERTKYKMKWMSKLYKP---INGKMYLLKLSKLVYKDMKI 569
Db 497 QPCPYGVKVVNNGSSNMEBEKN--GKCKSGLYPKPKDQBTGTYLLKSGGHND--- 552
QY 570 LKKNKEFCLTQSSSDGSVGYTTGASGNSKRLYDMCKYKHEVQKVVNGVEVE 629
Db 553 IEKLNKFCDEKNRGDTINSGGSGGSGNSGRLYEEKCYKQEDVYKGHNDDEE 612
QY 630 DDELKAGAGLCILPNPKNKEVESEKSONNHADIDKTFHDFEYVY/AHMLKDSIMPTK 689
Db 613 DYENKAGAGLCILKQKKNKEBEGGNTSEKPEIDOKTNPFFYVYAHMLKDSIMTK-K 671
QY 690 RLKSCISDKTYKC--RNGCNKCDCEKVVQKETEEMKPIDKHFKTJEGIPGEYFTLE 748
Db 672 KLQRCLONGNRJIKCGNKNKNNDCECFKRWITQKDEMGKIVQHFQNTIKRGSGDNTAE 731
QY 749 LI-----LKLQFLKED-----TEENTENSUDAEBABELKHLQKILKENENLAV 793
Db 732 LIPFDHYLYOYLDEFLKGSDESASEKSENSDAEABELKHLREIIESDNNQEAS 791
QY 794 VVAG--TECKTMDKLNHLELNATKCDPRLPEEDK----- 828
Db 792 VGGGTQKQINDKILNTEKDRADLCLEIHDEBEKEKGDGNECIJEBGENRYNCSGE 851
QY 829 ----- 828
Db 852 SGNKRYVLANKVAYOMHHKAKTQLASRAGSALRGDISLAQPKMIRNGSTLKGQICKIN 911
QY 829 -----SRGRSADPS-----PIIFTP----- 843
Db 912 ENYSNDSRGNSSGPGYKGDGAGVBMRIQTEWSNIEGRKQYSYKVFLLPRLREHMTSN 971
QY 844 ----- 843
Db 972 LENDVGSYTKNDKASHSLGVDYQLAAKTDAEIIKRYDQNNIQULTPYIQQKDEAMCR 1031
QY 844 ----- 843
Db 1032 AVRYSPADIGDILRGDMDEDEKSSITDMETRLITVFKNIKEKHGDIKDNPKYTGDESKKP 1091

QY 844 ----- 843
Db 1092 AYKRLRADMEANRHQVRAAMKCATKGIICPMYVDYIIPQRLRMWTBAMWYCAQSOE 1151
QY 844 ----- 843
Db 1152 YDLKKICADCMKSGDKCTOGDVDCGKACADCKYKEIEIKMNEQWKRISDKYNLYLQ 1211
QY 844 ----- 843
Db 1212 AKTSTNGRTVLGDDDDPYQOMVDFLTPHHKASTAARVLYKRAAGSTETATAAAPTTPY 1271
QY 844 ----- 843
Db 1272 STAAGYIHOEIGYGCBOEQOFCEKKHGATSTSTTKENKEYTFKOPPEYATACDINRS 1331
QY 844 ---RPEEKED----- 850
Db 1332 QTEBPKKKEENVESACKIYEKILEGKNGRTTVGBCNPKESYPMDCNNIDISHGACMP 1391
QY 851 ----- 863
Db 1392 PRROKLCIYTAHESQFENITDNLDAFIKTAAEFLSMOYTKSKNDSEAKILDNGL 1451
QY 864 ----- 875
Db 1452 IPSQFLASMYTFGDDYRDICLNTDISKQNDVAKAKDKIGKFESDKSPSGLSROEMW 1511
QY 876 ----- 879
Db 1512 KTNPELWKGMLCALTKYVYTDJDNKRRIKNDY SYDKVNOSONGNPSLEFAKPOFLRM 1571
QY 880 ----- 879
Db 1572 IEMBEERCAEROKKENITIKDACNEINSTQOCNDAKHRCNOACRAYOEYVENKKKEFSQT 1631
QY 880 ----- 879
Db 1632 NNFLKAVNOPQDPEYKGYEKGDVQPIQGNBYLLQCKDNKSCOMQNVLSVSPKEKF 1691
QY 880 ----- 903
Db 1692 GKVAHKYPEKDCYQGHVPSIPPPPPVQOPAPAPVTVVQVCSIV-KTEKDKDNNEFSDA 1750
QY 904 CSLAYGNNNSRLCWRVY---TPSEPTT---SSDKNGALICVPRRRRLYIKKIYDMATKT 957
Db 1751 CGLKY-GKTASSMKCIPSDTKSGAGATTGKSGDSGSICIPRRRLYVYKLDOWATL 1809
QY 958 ESPQASGEASSTSGSTTPPDSKELALKAFYESAIEFFLMHRYKEKKAVAQEGAGHG 1017
Db 1810 --PQEGGAAPSHSRA-----DDLRNAFIQSAIETFFLMDRYKEKKPQG-DGSQA 1858
QY 1018 LPRABE--GSEYEPDEK-LKEGKIPDQFLQMYTTLADYDILFSGSNDPTYSVKDPS 1074
Db 1859 LSOJLSTYSDEDEBPBKLLONGKIPDFELRMEYTLGDYDILVHGN--TSSGNTNG 1916
QY 1075 SSNDNLKILVLLASGSTEOREKKNKYKEI--KNBRKOSTERSAPNLVSHQPTWENNKG 1132
Db 1917 SNNN---NIVLEASGNKEDMOKIOEKIEQILPKNGCTPLVYKSS---AQPTDKMNEHAE 1970
QY 1133 YIMHWGVALT---SKDKIAKVEKKRPOKIEPENLME-----ANK-----PKP 1175
Db 1971 SIMGMICALITYETKPNPTDSARGDE--NKTEKDEYERKEFGSTAOKHGAASPTGYK 2027
QY 1176 POYOTYVAVKLDENGSTPRTTOTQASSDNPTTLTHFYKRPYTRWPEEMGESCREKK 1235
Db 2028 TOYDYEVKLEDTSG-----AKTPSASSDTP-LLDSEFLRPPRYLYEWMQNCCKRRKH 2081
QY 1236 RLKQIVDKYKENDVG-----RCSGGEACDSISTHYSVTPSPNGCGKHCSSY 1287
Db 2082 KLAQIKHECKVEENGSGSRKGITROYSGDEACNEMLPKMDGTVPLDEKSKPCSSY 2141
QY 1288 RKWIERKKIEFHKOSNAYGOOKTDATRNNGNTFDKEFCKTLETWPDAAKFLERLKNPCK 1347

Db 2142 RKWIESKKEFEKQEKAEQOK-DCVNGSNKHNGCEFTLTSSAKDFLKT--GPCK 2198
QY 1348 TNKEYGDDIDPEKDSKTFQTEYCGPCPKRTKQONQNGCVSLNGCDDKSIDAKEI 1407
Db 2199 PNNVEGKTIFFD--DDKTFKTKDCCDPLKFSVNCCKDECD-NSKGTDRKNNKIDATDI 2254
QY 1408 AKMRSSTTDVVMARYSDNDNTNFEGBDLKDACOHANIFKGIKRDYKMGCGYGVDCBOTN 1467
Db 2255 ENGVDSTV-LEMRYASBSKSGFNGDGLNACRAGIPEGIFKDKEMKCRNNGYVVCAPEN 2313
QY 1468 INERTDKEYIQIDALEKRWVENFLEDYKINDYISHCIRKBECSKICNGEKNKCLEK 1527
Db 2314 VNGEAKGKHIIQIDALYRWVEYEFEDYKIKHRIKSHIRKNGEISPCI---KN--CWEK 2367
QY 1528 WIEKKIAMENIKKRFNDQYENKQOPQYVKSILBELIPKAAVNDQDNVYKLCVFNESK 1587
Db 2368 WVDOKRKEWKETRFPODQYNDNSDDNVSFLETILPQITDANAKNKYIKLSKFGNSG 2427
QY 1588 GCTLISNTQ--NNKENDAIDMLKKGKAKNCPGK--PSGEKQSDCKEPPPLPDE--- 1639
Db 2428 GCSASANDQNKNGEYKDAIDCMLKKLKDKIGCEKKNHQTSDTECSDPQOPQTLDEFTLD 2487
QY 1640 ---EDQNEENTLEPPKFCPT--TQPEKGGETCGNKEEKDEKKEESEPAKEESGP 1694
Db 2488 DDIETEAKKNNM--PKICEVNLKTAQOEDEGC---C-----VPAENSEEPATDSGK 2535
QY 1695 AAERPAPAESEETNPPRPBGTPAPAPSTAP---PPDPPTPLRPQADEPFDSTIL 1751
Db 2536 ETPEOTPVLPKPEEAVPPEPP-----PQEKAPAPIQOPQPTPQILDHNVLTAL 2599
QY 1752 QT-TIPFGVALATSIAPFLPKKRTKASVGNLPQILQIPKSDYDIPRLKSSNRYIPVSD 1810
Db 2590 VTSILAMSVGIGFATFYFYIAKKTKSQVGNLPQILQIPKSDYDIPRLKSSNRYIPTSG 2649
QY 1811 RYKGRYIYMEGSDDEKRYAFASDPTDVTSSSEYEELINDIYVPSPKYKTLIEVLE 1870
Db 2650 KYRGRKYIYEGSDGJDS-GYTDHYSIDTSSSEYEEMDINDIYVPSPKYKTLIEVLE 2708
QY 1871 P-----SGNNTASGRNTPSDRNQINDGIPSSKITNENWQLKKEFTSN 1916
Db 2709 PSGNNTASGNNTASGNNTASGNTPSDONDINDGIPSSKITNENWQLKDEFISQ 2768
QY 1917 MLONOPNDPNDYTSNGNSTNTITTSRHNVDNNTWTMSROMMEENLLPSIHDCNLY 1976
Db 2769 YLOSEPNTERN-----MLGYVNDKNTPTSHHANEKPRFIMSHDNLNLF 2813
QY 1977 SGEYSYNV-----NMVN-----SMNDIPINRDNNVSGIDLINDLSGSKPI 2019
Db 2814 SGEENYDMFSGNPNPISIDSTNSMDSLTNNHSPYNDKNDLGSIDLINDALSNGH-I 2872
QY 2020 DIYDEVLRKRENELFGE-NKRTSTONVAKTTMSDPIHNLLEFHWLDOHRMCEKW 2078
Db 2873 DIYDEMUKRKRENELEFGKHNKHTNTYNAKPARDDPTQOINLFHMLDOHRMCEKW 2932
QY 2079 NKEDILNKLEEMKKNINNSGKTYNSDNKPSHNVLNTDVSIOIDMDNPKTKIEITNM 2138
Db 2933 NNHERLRLKLELM--ENETHSGDI--NSGIPSGHNVLNTDVSIOIDMDNPKTKIEITNM 2988
QY 2139 PTMKNHIEKNIVN 2210
Db 3049 PTMKNHIEKNIVN 3060
RESULT 3
US-08-568-459A-12
; Sequence 12, Application US/08568459A
; Patent No. 5849306
; GENERAL INFORMATION:


```

Db 1394 PRROKCLLYIAHESQTEINIKTDMLKDAFTKTAAEFLSMOYKKSNDSEAKTILDRGL 1453
QY 864 -----DEFAKE-----TTEGS----- 875
Db 1454 IPSQFLBSMTTFGDIYRICLNTDISKQNDYAKAKDKIGFFSKDGSKSPSGLSROBWM 1513
QY 876 -----ATDT----- 879
Db 1514 KTNPEIWKGMICALTKYVTDITDNKRKIKNDYSYDKVQNGQNGNPSLEPAKQFLRMM 1573
QY 880 ----- 879
Db 1574 IEMGEFCAEROKKENIKDACAINEINSTOQCNDAKHCNOACRAYOEYVENKKKEFGSGOT 1633
QY 880 ----- 879
Db 1634 NNFVLKANVOPDEPKYKYEKQDVOPLOQNEVYLQCKDNKCSMDGNVLSVSPKEKPF 1693
QY 880 -----TSLDVCPIYGVKLTIKDNESLQDA 903
Db 1694 GKVAHKYPEKCDYQKHKVPSIRPPRPVQPPPEAPYTVVOCSTIV-KTLFKDTNNFSDA 1752
QY 904 CSLKTYGKNSRLGRV---TPSGEPTT---SSDKNGAICVPRRRRLYIKKIYDMATKT 957
Db 1753 CGLYV-GKTAPSSMKCIPSDTKSGGATGKSGSDSGSICIPRRRLRYVGLQEWATLAL 1811
QY 958 ESPASGSEASTSGSTTPPSKELKALFAVESAIETFEFLMHYKKEKKAVADEGAGHG 1017
Db 1812 --PQGEBAAPHSHRA-----DRLRNAFTQSAIETFEFLMDRKEKKEKQOG-DGSOQA 1860
QY 1018 LPRVBE--GSPEDYDEDK-LKEGKIIPDGLRQMEYTLGDYBILFSGSNDTTSYKSDTPS 1074
Db 1861 LSQLSTSVDEDEBPPKLLQNGKIPPFLLMFTYTLGDYBILVHGN--TDSQNGING 1918
QY 1075 SSNDNLKNIYVLASGSTEOERKNNKYEI--KNFRKCTERSAPNLVSHPOTWENNGK 1132
Db 1919 SNNN---NIVLEASGNKEDMOKIOEKIEOILPKNGGTPLVKSS--AQPDKWMNEHAE 1972
QY 1133 YIWMGVNICALT---SKDKIAKGYEKKPKOKTENPENLMD-----ANKK-----PKP 1175
Db 1973 SIMGMICALTYTEKNEPDTKSARGE--NKTEKDEVEYEKFGSTADHAGSTPPTGYK 2029
QY 1176 POYOYTNVKLDENGSGTSPTTQTOASSDNTPTTLTHFVKRPTFRMEWEGESFCREK 1235
Db 2030 TQYDYEKVKLEDTSG-----AKTPSASDTP-LISDFYLRPRYRILEMONGNCKKRXH 2083
QY 1236 RLKQIKYDCKYENGDVG-----RCSGDGEACDSISTHDYSTVPSPFCGCGKHCSSY 1287
Db 2084 KLAQIKHECKVEENGSGSRGKITROYSGDEACNEMLPKNDGTVPDLKPSCAKPCSSY 2143
QY 1288 RKWIERKKIEFKHOSNAYGOOKTDATRNNGTPEKFCCKTLETMPDAKFLERLKNCK 1347
Db 2144 RKWIESGKEEKEKOEKAYEOK-DKCVANGSKHNGFCELTITSSKADFKTL--GPKC 2200
QY 1348 TNKEYGDDIDFEKDKTFQTEYCGPCPKFKTNCQNGMCGVSLNGNCDKSDIAKEI 1407
Db 2201 PNVEGKTIFF--DKTFTKTKDCDCLKFSVCKKOEDC-NSKGTDRKKNIDATDI 2256
QY 1408 AKMSSSTTDVVMARYSDNDNTTFEBDDLKACQAHANIRKIGIKDYWKCGYCVGVDICEQTN 1467
Db 2257 ENGVDSTV-LEMRVSAKSGFNGDGLNMACRGAGIEGIRKDKMCKGANNCGYVYCKREN 2315
QY 1468 INERTDGEYIOIRALPKRWENFLEDYKINDKISHCIRKGEKSGKINGCKSKCKEK 1527
Db 2316 VNGAKKHHIIQIRALPKRWYEFEDYKNTKIKHISHRIKNGELSPCL---KN--CUEK 2369
QY 1528 WIEKKIAEMENIKRFDYQENKQDOPYNKSILEELIPKIAVVDNDQNVIKLCVFENSK 1587
Db 2370 WVDCKRKMKTITERFDYQNDSDDNVRSFLETILPQITDANAKNKVIKLSKFGNSC 2429
QY 1588 GCTLISTNQ--NKNENDADCMKKLGKAKNCPGK---PSGEGKQSDCKEPPPLPDE--- 1639

```

```

Db 2430 GCSASANEQNNKNGEKDAIDCMKKLKDRTGECEKHHQTSDETCSDTPQPOPULEDTLD 2489
QY 1640 ---EDONPEENTLEPPKPCPT--TOPPEKGGFTGCKNEKKEKKESEEPAKESGP 1694
Db 2490 DDIETEEKKMM--PKICEVNLKTAQOEDBG---C-----VPAENSEPAPATBSGK 2537
QY 1695 AAEPAPAESEETETNEPPEPGTGPAPSTPAP---PTPDTPPPLRPOADEPFDSTIL 1751
Db 2538 ETPROTPLYKPEEAVPEPPPP-----PPOEKAPAPIQPOPPTPTQLLDNPHVLTAL 2591
QY 1752 QT-TIPEGVALALGSIAPFLK 1772
Db 2592 VTSLAMSVGIGFATFYFLK 2613

```

RESULT 4

```

US-08-487-826B-12
; Sequence 12, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Welliams, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelson, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CPI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2710 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
US-08-487-826B-12

```

Query Match 26.4%; Score 3197.5; DB 2; Length 2710;

Best Local Similarity 30.6%; Pred. No. 1e-203; Matches 834; Conservative 244; Mismatches 569; Indels 1075; Gaps 74;

```

QY 4 SCGSGGQDDEDAKVVDEFOGKVDVEHGEAKNVVSLKSLASLIGTAFTVKSMT 63
Db 14 AAGGDDIEDSAKHMFRIGKDVYDKVKEKRGKGLQRLS-----EAKFEKNESDP 67

```

Qy	64	ES-----	KYTELLEANSKRPPC-----	KQKGNVDVRFSEYKQADGNKKMKC	107
Db	68	QTPEDPDLBHKHTNTT-TN	VINPCADRSVDVRFSEYGGQSTHNHRKCSQOQDNGK---	123	
Qy	108	SNGTQCAPRRLHLCNKNFPMNMS	NDSSKAKHDLLEAVCAAAUYEGESITHTYPRKDSKY	167	
Db	124	----ACAPYRLHACDONLEOIE	RIKTTN--HNLDYVCAAAFEBOGSTITQDYPKQATY	178	
Qy	168	PGSDPFCMTMLARSFADIGDI	IRGRDYLIGNKKKONGKQETEREKLEQKLEJFKKIHDN	227	
Db	179	GDSPQJCTMLARSFADIGDI	VRGRDYLGNPQEIK-----QNOOLENNKTIJFKIYER	233	
Qy	228	LKQKEAQRRNGDEDPNFYLR	EDPMWTANFEYWGANTCSKEIDNSSYFATGNDTGQGE	287	
Db	234	LNGAEA--RYG--NDPEFFKL	REDMWTANFEYWKATCNAM--GNIFYHATON---RG-	283	
Qy	288	SQTHNRCORCKDKGANGKRA	GGQDVITYPTFDYVQYLRJMFPEABDFCKKKKKKLE	347	
Db	284	ERTYGYCRDNDQ-----	-PPTYFDYVQYLRJMFPEABDFCKKKKKKIK	327	
Qy	348	NLEKQSGKDKSDEBYRYCS	RHNGYDCEOPTISRKCKVRMGKCTQCFPACSGEUMJDNOK	407	
Db	328	DYKNCNGKDKEDDRICS	NRNGYDECTKRAIKLRGKCISGLACNRYVWINQKE	387	
Qy	408	QEDKO-KKYTKETI-----	SDGGGRKKRAVVG--TRYEGEYSFYEKLKNDGYGVDAE	458	
Db	388	QFDQKKKKYDEBEIKEN	YGASGSGRQKRDAGTJTTYNDYEKKFYEDELNKSSEYRDXE	447	
Qy	459	LGLLNKCAKCDITD--GKI	INFEVNSGGVYVGGSGGSGSGSNTDENKGFYFSEYC	516	
Db	448	LEKLSNEICTKRVKDEGG	ITDFKNV-----SDTSGASGYNVESQGTFRYSKYC	498	
Qy	517	QPCDCEVO--HKGG--	NQMERKTKYVKMMSKLYP---INKMYVLISLKVUDMMI	569	
Db	499	QPCYGVCAKYNNGSS	NEBDEKKN-CCKASGKIYERPKBEGTJTITLKSQGHND---	554	
Qy	570	LKKMKEPCLTONSSDGS	VSGSVYTTGASGNSSEKLEIYEMKCYKHNEDQYVNOVEEBE	629	
Db	555	IEEKLNKFCDEKKN	DNJINSGGSGSGSGSGSQNOELIYEMKCYKGEDVYVGHNDEDEE	614	
Qy	630	DDDLKAGGCLTLPNKK	KKVESEAKSONNHADIOKTFHDFYVYVANHKLKOSTHWRTK	669	
Db	615	DYENAKNAGGLCTILK	KNKKNEBEGNTSEKPEBIOKTFHDFYVYVANHKLKOSTHWK-X	673	
Qy	690	RLKSCISDGTAKMC--	RNGCNKKCDCEFKWYKOKETEMKPKIDHFKTQOEGREGYFUTLE	748	
Db	674	KLOKCLONGNRKIKG	NNKKNKCNDECEFRMYITQOKDEGKIYQVHFQYONIKRGSGSDNTAE	733	
Qy	749	LI-----	LKLOFLKED--TEENTENSLDAEAEELKHLOKILKENENNLA	793	
Db	734	LIPEDHUYLOYNLQ	EBEFLGDSDBASEKSENSLDAEAEELKHILREIIESDNNQEAS	793	
Qy	794	VNAG--TEQKTLMDKIL	NHELNDATKCDCLPREDK-----	828	
Db	794	VGGGVTEOKNIMDKL	NYEKEDEADLCLEIHDEBEEKGEKGDNECIEGENFRYNPCSGE	853	
Qy	829	-----	-----	828	
Db	854	SGNKRYEVLANKVAYO	NNHAKTQYLASRAGBSALRDISLAQFRKNGANGSTLKGQIOKIN	913	
Qy	829	-----	-----	843	
Db	914	ENYSNDSKNGSGG	PTCKDDHGCVRMRICTGEMSNIEGKKQOTSYKNVFLRPRRHHOTSN	973	
Qy	844	-----	-----	843	
Db	974	LENLDVGSVTKNDKASH	SLLDGVOLAAKTDAELIKRYKDONNIOLTDPYQOKDOEAMCR	1033	
Qy	844	-----	-----	843	
Db	1034	AVRISPADLGDITING	ROMBDEDSSTJMETRILITVFENIKENKIDGKDNPKYTGDESKRP	1099	
Qy	844	-----	-----	843	

Dd	1094	AYKKLRADMEANRHOYWRAMKCATKSI	ICPGHVPUDYI	IPOLRMATE	YAEWYCAQSOE	1135			
OY	844	----	----	----	----	843			
Dd	1154	YDKLKICADOMSGDGCKTQGVVDGCKCAACDXYKKEEIKENBQWKRISDKYMLYLQ				1213			
OY	844	----	----	----	----	843			
Dd	1214	AKTTSTNPGRTVLGDPPDDPYOQWVFLTPILHKASIAARYLVKRAAGSPYEIAAAPTTPY				1273			
OY	844	----	----	----	----	843			
Dd	1274	STAAGYIHQELGYGCOEQIOFCEBKHGATSTYTEENKEYTEFPKPPFEYATACDCINRS				1333			
OY	844	----	----	----	----	850			
Dd	1334	QTEBPKKKEENVESACKIVERKILEGKNNGRTTVEGCPNPKSYPDMDCN	NIDISHGACMP			1392			
OY	851	----	----	----	----	863			
Dd	1394	PRROKCLLYYIAHESOTENIKRTDNLKDAEIKTAAETFLSMOYKYS	NDSEAKILDRGL			1453			
OY	864	----	----	----	----	875			
Dd	1454	IPSOFLSMATTFPDYDIDLNTDISKONDAKAKDKIGKFFPSKDGK	KSPBGLSRQEW			1513			
OY	876	----	----	----	----	879			
Dd	1514	KTNGBELWKGMLCATRYVYDTONKRRKIKNDY	SYDRVNSQNGNPBJ	BEFAKPOFLRM		1572			
OY	880	----	----	----	----	879			
Dd	1574	IEWGEFCABEROKKENTIKDACEINSTOOCNDAKHRCNOACRAYOE	YENKKKEPSGOT			1633			
OY	880	----	----	----	----	879			
Dd	1634	NNFLKANNVQPODEYGYEYKQGVQPIQENELLOKONNKCSCMD	NVLVSAPKEKPF			1693			
OY	880	----	----	----	----	903			
Dd	1694	GKYAHKYPEKCDQYGHKVPISIPRPPPVYQPOEAPETVYDVCSIV	KTLFMDTNNFSDA			1752			
OY	904	CSLYVGGNNSRLGRCY	----	TPSGEPTT	----	SSDKGATCVPRRRLYIKKIVMATPT	957		
Dd	1753	GLXY	GRTAASSKCC	IPSDPTKSAGATYTKSSDSGSIC	IPRRRLVGVGLQENATL	1811			
OY	958	ESPDSASSEASTSGSTTTPDPSKEALKAFAVESAAIETFFLMHRYEERKAAVQACAGHC	1017						
Dd	1812	----	----	----	----	1866			
OY	1018	LPRVEE	GSPEYDPEDK	LKEGKIPDGLRQWFFYLGDY	RDILSGSNDTTSVSKDTPS	1074			
Dd	1861	LSOLTSTYSDEEDBPDKLLONGKIPDPFLRLAFYTLGDI	RDILVHGN	----	----	1918			
OY	1075	SSNONLKIVYLASSGTEOEERKANKKKEI	----	GNFKRCSSTERSAPML	VSHPOTWENNKG	1133			
Dd	1919	SNNN	----	NIVLEASGNKEDMOKIOEKLEO	LLPKNGGTPLYPKSS	----	AOTPKMNEHRE	1972	
OY	1133	YIWHGMVCAIT	----	SKDKIAGKVEKKPKOTIENPMLMDE	----	AKRK	----	PKP	1175
Dd	1973	SIMGMICALYITEKKNPDTARSARGE	----	NKIEKDEYVEKEFPGSTAL	KHGTASTPTGYK	2022			
OY	1176	POYOTYANKLDENSGTSPRTTQTOASSDNTPTTLHFVAKRPTFYFM	IEWGSEFCERKK	1233					
Dd	2030	TOYDERKVKLEDTSG	----	AKTPSASDTP	LLSDFVILPRPYFRY	IEWGCONFCFKRRH	2083		
OY	1236	RLKIOIKVDCKEVENDVG	----	RCSSGGEACDSISTHDSIVF	FJNCPCGKHCSSY	1287			
Dd	2084	KLAQIKHECKVEENGSGSRGGITROYSGGECANEMLRKNDGTYH	LEKPBSCAPCSSY	2143					
OY	1288	RKWIERKKEIFEHKOSNAYGOOKTDATFNNNGTPEDKCEFKTLETWPD	IAFLERLKNGPCK	1347					

Db 2144 KMWIESKGEFEKOEKAEVOK -DKVNGSKHNDNGCELTLYTSSKADFLKT--GPCK 2200
Qy 1348 TNKEYGDDIDPEKDSKTFQHTCYGCPKFKRTNCGNCGVSLNGCNDKSDAKEI 1407
Db 2201 PNNNEGKTFD---DDTKFKTKDCDCLKFSVCKKDECD-NSKGTDCRKNKSIDATDI 2256
Qy 1408 AKMRSSTDDVMRVSDNTFTESDDLKDACQAHNIFKGIKRDVWKGCVGVDICEQTN 1467
Db 2257 ENGVDSTV-LEMRYASDSKSGFNGDGLFNACRGAGIEGIRKDKMCKGNVGVYCKPEN 2315
Qy 1468 INERTDGEYQIRALKRWRWENFLEYNKINDKISHIKKDESKCINGEKSKCLEK 1527
Db 2316 VNGAKGHHIQRALKRWVEYFEDYNTKHKISHIRIKNGEISPC1---KN-CVEK 2369
Qy 1528 WIEKRIAWENIKRRFNDQYENKQDPYVKSILELPIKJAVVNDODNVIKLCVFENSK 1587
Db 2370 WVDKREKREKITERFDQYKNDSDDDNVASFLETILPQITDANAKNKVTKLKFNGSC 2429
Qy 1588 GCTLIISNQ--NKNENDIDMLKLGKAKNCPK--PSGEKQSDCKEPPPLPDE-- 1639
Db 2430 GCSASANEQNGEYKDAIDMLKKDKICECEKHHQTSDECSPTPQPTLEDETLD 2489
Qy 1640 ---EDQNPENTLEPPKFCPT--TOPPEKGGETCGNKEKDEKKESEEPKESGP 1694
Db 2490 DDIETEAKKNM--PKICENVLKTAAQEDEG--C-----VPAENSEEPATDSGK 2537
Qy 1695 AAEPAATAESEETETNPPEPTGPAAPSTPAP---PRPDTPPLRPQADEPDSITL 1751
Db 2538 ETPEQTVLKPEEAVPEPPPP-----PROEKAPAPLPQPTPQOLDNHVLTA 2591
Qy 1752 QT-TIPGVALAGSIAFLK 1772
Db 2592 VTSLAVSGIGFATFYFLK 2613

RESULT 5
US-08-568-459A-10
; Sequence 10, Application US/08568459A
; Patent No. 5849306
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knodde Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; Zip: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/568,459A
; FILING DATE: 07-DEC-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israel, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121, 001CPI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 10:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 700 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
; US-08-568-459A-10

Query Match 11.4%; Score 1385; DB 2; Length 700;
Best Local Similarity 40.8%; Pred. No. 5,5e-84;
Matches 328; Conservative 106; Mismatches 207; Indels 162; Gaps 34;

Db 113 CAPRRHLCKNPPMNSNDSSAKHDLAEVMAKYGESIKTYPKYDSYPSDF 172
Qy 10 CAPRRHLCDY---NSESIDTSTHKLLEVCMAKYGESINTHYTOHQTNEEDSAS 66
Db 173 PMCTMLARSPADIGDIIRGRDLYG--NKKKKONGKETEERKLEOKLKEIFKTIH-DNLK 229
Qy 67 QLTVALARSPADIGDIYRGKDLVGYDNKEKEQ-----RKLQKLKDFKTKHKVYMK 120
Db 230 DKEAKRYNGD-EDNPFYKLEDMWTANRETVMGAMTCSKELDSSYFRATCNDTGGPS 288
Qy 121 TNGAQERYIDAKAGDFEQLRDMWTSNRETVMKALICHAPKEANYFTKACN-VGKG-- 177
Db 289 QTHNKKCRDKDKANAKPKRAGDVTITVPIFYVQVQLRMEFEAMDEDCRKKKKLEN 348
Qy 178 -TNGQCHC-----TGGD-----VPIFYDLYQVLRMEFEAMDEDCRKKKKLEN 220
Db 349 LEKCRGKDKSDEYRYCSRNGYDEOTISRKGVRMGCGTDCDFACGSYENWIDNORK 408
Qy 221 LQKOCRDYEDN---LYSGNGYDCTKTIYKKKGLVIEHCTNGSCWCMETWIDNCKE 277
Db 409 FDKOK-KYREISDGG-----RRRAVGTTKYE--GYEKSEYKLNQDGYTDAFLG 460
Qy 278 FLKOKRKYETIEISGSGSKSPKRTKRAARSSSSSDNGYSEKFKLKEVYQDQVDFLK 337
Db 461 LLNNEKACKDTDGGKINFEVNSGGGVGSGTSGASTNDENKGTYSRSYCOQPC 520
Qy 338 ILNKEGICQKQPOVG--NEKADN-----VDFTNKLYKTSRTEICEPCP 380
Db 521 DCGVHKGQGWEMERTYKKKRWMSKLYKPIKNGVLLKSLKLVKDKMILKKMKKECLT 580
Qy 381 WGLIE-KGGRPW--KVAGDKTCGSAKTYDPKNTIDIPVLPRKSOONILKKYKNC-- 435
Qy 581 QNSSDGSVSVYTTGASGNSSEKKELYDEMKCYKHNEVQKVNVOGEVEEDDELKAGGL 640
Db 436 -----EKGAPGGQIIRK-----WQCYV-----DEHR----- 456
Qy 641 CILPKNKKNVSPAK-----SQNNHADIQKTFHDFEYVVAHMLKDSIHRTRKLSKI- 695
Db 457 ---PSSKNNNNCEGTWDTKFTQKQT--VKSYNFEFMDWMDMLHDSVEWKTE-LSKGIN 510
Qy 696 --SDGKTMKCRNGCNKKKDCFEKAVKOKETEMKPIKDFHTQEG1-----PRGY 743
Db 511 NNTNGNTRNNNNKCKTCCGCFQKVKVEKKQDEMAIKHFKQNDIYQOKGLIYSPYG-- 568
Qy 744 FTYLELILK---LQLEKDETEENTENSLEAEAEELKHLQKILKENNNLAVNAGTE 799
Db 569 --VYDLVLKGNLLQNIK-VYHGDY-----DDIKHITKL--LDEEDAVAVVLGSKD 614
Qy 800 QKTLMDKLHMLNDATKCK---DCPLPREDSRGSAPSPDIFLRP-----E 846
Db 615 NTYI-DKLQHEKQAOQCKOKODECEKKAQOESRGSALFREDERTQOPADSAGEVEE 673
Qy 847 EKEDENEDDEDEVRDEETAK 869
Db 674 EDDDDYDEDDDDYVQDVAYSE 696

```
RESULT
6
US-08-487-826B-10
; Sequence 10, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 700 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
US-08-487-826B-10

Query Match 11.4%, Score 1385; DB 2; Length 700;
Best Local Similarity 40.8%; Pred. No. 5.5e-84;
Matches 328; Conservative 106; Mismatches 207; Indels 162; Gaps 34;

QY 113 CAPRRLLICNKNFNMNSNDSSKAKHDLIAVCAAKYEESITHPKYISKTPGSPF 172
DB 10 CAPRRLLICVD--NLSSIDTSTTHKLLLEVCAAKYEGNSIMPTTHQHTNEDSS 66
QY 173 PWCTMARSFADIGDIINGRDLYLG--NKKKRONKETEREKLEQKLEIFKIH-DNLK 229
DB 67 QLCYLAASFADIGDIVGKDLVGYDNKEKQ-----RKLLEQKLDIFKIKHDKVAK 120
QY 230 DKEAKRYNGD-EDPNFYKLRDWMPTANRETVGAMTCSKELDNSSYFPRATCNDTGQSPS 288
DB 121 TNGAERYIDAKGDFQLREDWMTSNRETVWAKALICHAPKEANYFIKTACN-VGKG-- 177
QY 289 QTHNRCRCDKDKGANAGRPKAGDGVTVTPYTFEDVVPQYLRFEERAEDEFCRKKKKLLEN 348
DB 178 -TNGCCHC-----IGD-----VPTTFEDVVPQYLRFEERAEDEFCRKKKKLLEN 220
QY 349 LEKQCRGKDKSDERYCGRNGYDCQOTISRKGRVMKSGKCTDPCFACSGSEYMNINORKO 408
DB 221 LQKQGRDVEQN---LYCGNGIDCTKITIKKQKLVIGEHCTNCSYWCNRYETWINDQKKE 277
```

```
QY 409 FDKQK-KYTKETSDCG-----RKRRAVGCTTKYE--GYEKSFYEKLMDGYGVDAFIG 460
DB 278 FLKQRRKYETEISGGSGSKSPKRRRAARRSSSSSDNGYESKFRKLKLVGYQDDVDFLK 337
QY 461 LLNNEKACKDITDGGKINFEKNSGGVGVGGSGGTSAGSANDENKKGFFYSEXCOPCP 520
DB 338 IINKEGICOKOPVG--NEKADN-----VDFNEXYVKTFSTETCEPCP 380
QY 521 DGVQVHKGQNMERTKYKRMRSKLYRPINGKAVLLKSLKLVYDMMILKKNKEPCLT 580
DB 381 WGLE-KGSPW--KYKGDKTGSAKTKTYDPKNTDIPVLVLPDRSQYLKKYNFC-- 435
QY 581 QNSGSGVGSVYTTGASGNSSEKELIDEMKCYKANEVQKVVQGEVE-SDDELGAGGL 640
DB 436 -----EKGAPGGQIRK-----WQCY-----DEHR----- 456
QY 641 CILPMPKKNKEVSEAK-----SONNHADIOKTFHDPFFYVAHMLKDSIIMPTRLKSCI- 695
DB 457 ---PSSKNNKNCVEGTMDKFTGKOT--VKSYNVEFWMDVHMLDSDVEMKTE-LSKCI 510
QY 696 --SDGTYMKCRNGCNKKKDCFEKWKVQKETEKKPIKDEKTOEGT-----PEGY 743
DB 511 NNTNGNTCANNKKCKTDCGCFQKWKVEKKQOEWMAIKDHEGKQTDIVQKGLVFSPYG-- 568
QY 744 FTTLELIILK----LQFLKEDTEENTENSIDAEAEELKHLQIILKLENENNLAVYNAGTE 799
DB 569 --VLDLVYLGKMLQNIK-DVHGD-----DDIKIKKL--LDEEDAVAVVVLGGKD 614
QY 800 QKTLMDKLLNHELDNATCK---DCPLPEEDKSGRSADPSPDIFIPRP-----E 846
DB 615 NNTI-DKLLQHEKEQAEQCKQKQECCEKKAQOESGRSARETREDERTQCPADUSAGEVEE 673
QY 847 EKEDDEDDEDEVRRDEETAK 869
DB 674 EDDDDYDEDEDDEDVQVDVASE 696

RESULT
7
US-08-568-459A-8
; Sequence 8, Application US/08568459A
; Patent No. 5849306
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/568,459A
; FILING DATE: 07-DEC-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CP1
; TELECOMMUNICATION INFORMATION:
```



```

QY 835 DPEIDIFPREKEEDDEDDEDEVDDEDEAKETTESATDTTSLDVCPIYGVKVL 894
Db 463 -----DKN-----SVPTNKV-----473
QY 895 KDNESLADACSLKYGNNRLGMRCVPSGPTSSDKNGAICVPRRR-----RLY 946
Db 474 -----WEC-----KNPYLSTKD--VCVPRRQELCLGNIDRIY 505
QY 947 IKKIVDAKTESPOASGEASSTSGSTTPPSKELLKAVESNAI-ETFFLHRYKEE 1005
Db 506 DKNLL-----MKHEILALAIYESHLLKRYKKN 534
QY 1006 KKAQAGAGHGLPRVEGSEPEYDEPKLKEGKIPDGLRQMFYTLGDRILFGSNDT 1065
Db 535 -----DOK-----EYCKIINTFADIRIIT--GGTDY 559
QY 1066 TYSKDTPPSSNDNLKNIIVLASGSTEOERKMKYKEIKFRKCSSTERSAPNLVSHPT 1125
Db 560 W-----NDLSNRKLLVGKINTNSKYVHRNK-KNDKLFK-----DE 592
QY 1126 WMENNGYIMHGWCALTSKOKIAKVEKKRQKTEPENLMDANKKPKPQYQYTNKL 1185
Db 593 WKKYIKKDVWN--VISWPKDQV--CKEDDIEN-----622
QY 1186 DENSGSPRTTQTOASSDNTPTTLHFVKRPTVPRWFEWGESFCRERKKRLOIKYDCK 1245
Db 623 -----IPQFRWFSEMDDDYCODKTKMIEFLKVCCK 653
QY 1246 VENGDCVRCSGDGACOSISTHYDSTVPSFNCPCGKHCSYRMYIERKKTIEPKOSNAY 1305
Db 654 EK-----PCBED-----NCKSKNSYKEMISKKEEYKNOAKOY 687
QY 1306 GOKTDTARNNGNCFDEKFCLETPWDAKFLRLKKNPCCKTKVEYGGDDIDEEKSKT 1365
Db 688 QEYO-----KGNNTMYSEF-KSIRPEYLLKYSKCSN-----LNEDEKKE 729
QY 1366 FOHTEY--GCPCKFR-----TNCQNGCGVSLNGN-----CDDKSIDA 1404
Db 730 ELHSDYKNCJMCPEYVDVPSIIRNNQTSQEAVPRENTFIAHRTTPSISEPKMGEO 789
QY 1405 KEIKKMSSTTDVYMRVSDNDNTFEGDDLKDACOHANIEFKIRKDY--WKGVCYGVDI 1462
Db 790 KE-----RDDDSLKISVSPENSRPEIDAKDT--SWLTK-LKGVDISMPKAVIGSSP 839
QY 1463 CEQTNINERTDGKEYIDIRALFKRWENFLEDYKINDKISHCKKKGSGKINGCEKNS 1522
Db 840 NDNINATEQDN-----ISGV--NS 857
QY 1523 KCLEKMTIEKIAEWENIKRRENDQYENKDDPYNVKSILEELIPKIAVNDQDVNIKLV 1582
Db 858 KPLSDVDRP-----KKELEDQ--NSDESEETVWN--HISKPSINNOD-----897
QY 1583 FENSKGCTLSINTONKNENDAIID-----CMLKKLVKAKNCPGKSGEKOSDCKEPPP 1635
Db 898 -DSGSGSATVSESSSSNTGLSIDDRNGDIFVRTQDANFEDVIRKENADKDEDEK--953
QY 1636 LPDEODNPENITLPPKFCPTTOPPEEK-----GGTCGNKEKKKDEKKESEPERA-1668
Db 954 -ADERHSTSE-----SLSSPECKMLJNEGNSLNHBEVKEHTSNSNVQOSG 1001
QY 1689 -----KEESGPAABEPAPTAASEETETNEPEPPGTGPAAPSTPAPPTDTPPLRPQ 1741
Db 1002 GIYVANNVEKELKDTLENPSSLDGKAHELESPNLSSDOOMSTPG-----1048
QY 1742 ADEPFDSSTIIQTTIPFGVALAGSIAFLKKTAKYAVGNLFQILQIPKSDYDIPTLKSS 1801
Db 1049 -----PLDNTSEETTERI-----SN 1063
QY 1802 NRYIPYSDRKKGKTYIYMEGSDDEKAYAFMSDTTDTVSSSEYEELDINDIYPPGSPKY 1861
Db 1064 NEY--KVNER-----EDERTLTKEVEDIVLASHMNRRESDDGELY-----1100

```

```

QY 1862 KTLIEVLEPSGNNTTASGKNTPSDTRNDIONDGIPISSKITDNEQMLKKEFISNMLQ 1921
Db 1101 -----DENSDLSLVNDESEDA--EAKMKGN-----TSEMSHNS 1132
QY 1922 PNDVPNDYTSNGSSTNTNTTTSRHNVDNNTTMSRDNMEENLLPSIHGNIYSGBEY 1961
Db 1133 SQHIESDOQKNDKMTYVGLGTT--HVONEISVPTGEBDEK--LRESKSKYHKAEE 1186
QY 1962 SYNVNMYNSMNDIPINRNDNVYSGIDLINDLSGCKPIDIYDEVLLKRENELFCTENTKR 2041
Db 1187 RLSHTFDIKIN--PEDRSNTLHLKDIRNE-----ENBRH 1219
QY 2042 TSTQNV-----AKTNSDPINHOELFHKWLDRRHDCMKKKNEDILNK 2086
Db 1220 LTQNNINISQERDLQKHGFHTMNLHGDGYERSQIHN--SHHGNNQDRGNSGNVLNM 1276
QY 2087 LKEEMKENINNSGKTTNSDNKPSHNVLNTDVSIOQIDMDNPKTKNETTMMDTQDXSTM 2146
Db 1277 RS--NNNNFNIPSRNYLYDK-----KLDLDYENRNDSTYKELIKLAEI 1320
QY 2147 DTILDDLE-KYND 2158
Db 1321 NKCENEISVKICD 1333

```

```

RESULT 10
US-08-487-826B-4
; Sequence 4, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Slim, Kim L.
; APPLICANT: Chluis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; Zip: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487, 826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israeljen, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CPI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-0176
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1435 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
; US-08-487-826B-4

```



```

:
: APPLICANT: Miller, Louis H.
: APPLICANT: Peterson, David S.
: APPLICANT: Su, Xin-zhaun
: APPLICANT: Wellens, Thomas E.
: TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
: TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
: NUMBER OF SEQUENCES: 37
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Knobbe Martens Olson & Bear
: STREET: 620 Newport Center Drive 16th Floor
: CITY: Newport Beach
: STATE: California
: COUNTRY: US
: ZIP: 92660
: COMPUTER READABLE FORM:
: MEDIUM TYPE: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentln Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/568,459A
: FILING DATE: 07-DEC-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Israelsen, Ned
: REGISTRATION NUMBER: 29,655
: REFERENCE/DOCKET NUMBER: NIH121.001CP1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (619) 235-8550
: TELEFAX: (619) 235-0176
: INFORMATION FOR SEQ ID NO: 18:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 362 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: FRAGMENT TYPE: Internal
: ORIGINAL SOURCE:
:
: US-08-568-459A-18

```

```

Query Match          3.1%; Score 379.5; DB 2; Length 362;
Best Local Similarity 31.3%; Pred. No. 1,7e-17;
Matches 93; Conservative 14; Mismatches 149; Indels 41; Gaps 5;

```

```

QY 113 CAPPRRLHLCKNFPNMSNDSKAKHDLAEVCMARKYEGESIKTHYPKYDSKYPGSDF 172
    ||:|||||:
DB 2 CAPPRRLHLCDY---NLKXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 58
QY 173 PNCMTLARSFADIDIGDIRGRDLYLG--NKKRKQNGKETEREREKLEOKLEIFKKIHDNLKD 230
    :||:|||||:|||||:|||||:
DB 59 QLCIVLARSFADIDIGDIRGRDLYLGIDYDKKXXXXXXXXXXXXXXXXXXXXXXXXXXXX 118
QY 231 KEAKRKYNGDEDPNFYKLRDPMWTANRETVGAMTCSKELDNSSYFRATCDTGOGPSQT 290
    ||:|||||:|||||:|||||:
DB 119 XXXXXXXXGSD---FQRLREDMWTSNRETVMKALICHAXXXXXXXXXXXXXXC----- 164
QY 291 HNKCRCDKDKANAGKPRAGGDVTVPTFYDVPQYLRFWEEMAEDFCRRKKKKLEMLE 350
    ||:|||||:|||||:|||||:
DB 165 -----XXXXXXXXXXXXXXXXXXXXXXXXXXXXVPQYLRFWEEMAEDFCRRKKKKLEMLE 214
QY 351 KQCRGKDKSDERYRCNRNGYDCEOTISRKGKVRMGKCTDCCFACGSEYENWIDNQRK 407
    ||:|||||:|||||:|||||:
DB 215 KQ-----CXKXXXXXXXXXXXXXXXXXXXXCTNCSVMCRMYETWIDNQRK 259

```

```

RESULT 12
US-08-487-826B-30
; Sequence 30, Application US/08487826B
; Patent No. 5993827

```

```

:
: GENERAL INFORMATION:
: APPLICANT: Sim, Kim L.
: APPLICANT: Chitnis, Chetan
: APPLICANT: Miller, Louis H.
: APPLICANT: Peterson, David S.
: APPLICANT: Su, Xin-zhaun
: APPLICANT: Wellens, Thomas E.
: TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
: TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
: NUMBER OF SEQUENCES: 45
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Knobbe Martens Olson & Bear
: STREET: 620 Newport Center Drive 16th Floor
: CITY: Newport Beach
: STATE: California
: COUNTRY: US
: ZIP: 92660
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentln Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/487,826B
: FILING DATE: 10-SEP-1993
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Israelsen, Ned
: REGISTRATION NUMBER: 29,655
: REFERENCE/DOCKET NUMBER: NIH121.001CP1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (619) 235-8550
: TELEFAX: (619) 235-0176
: INFORMATION FOR SEQ ID NO: 30:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 362 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: FRAGMENT TYPE: Internal
: ORIGINAL SOURCE:
:
: US-08-487-826B-30

```

```

Query Match          3.1%; Score 379.5; DB 2; Length 362;
Best Local Similarity 31.3%; Pred. No. 1,7e-17;
Matches 93; Conservative 14; Mismatches 149; Indels 41; Gaps 5;

```

```

QY 113 CAPPRRLHLCKNFPNMSNDSKAKHDLAEVCMARKYEGESIKTHYPKYDSKYPGSDF 172
    ||:|||||:
DB 2 CAPPRRLHLCDY---NLKXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 58
QY 173 PNCMTLARSFADIDIGDIRGRDLYLG--NKKRKQNGKETEREREKLEOKLEIFKKIHDNLKD 230
    :||:|||||:|||||:|||||:
DB 59 QLCIVLARSFADIDIGDIRGRDLYLGIDYDKKXXXXXXXXXXXXXXXXXXXXXXXXXXXX 118
QY 231 KEAKRKYNGDEDPNFYKLRDPMWTANRETVGAMTCSKELDNSSYFRATCDTGOGPSQT 290
    ||:|||||:|||||:|||||:
DB 119 XXXXXXXXGSD---FQRLREDMWTSNRETVMKALICHAXXXXXXXXXXXXXXC----- 164
QY 291 HNKCRCDKDKANAGKPRAGGDVTVPTFYDVPQYLRFWEEMAEDFCRRKKKKLEMLE 350
    ||:|||||:|||||:|||||:
DB 165 -----XXXXXXXXXXXXXXXXXXXXXXXXXXXXVPQYLRFWEEMAEDFCRRKKKKLEMLE 214
QY 351 KQCRGKDKSDERYRCNRNGYDCEOTISRKGKVRMGKCTDCCFACGSEYENWIDNQRK 407
    ||:|||||:|||||:|||||:
DB 215 KQ-----CXKXXXXXXXXXXXXXXXXXXXXCTNCSVMCRMYETWIDNQRK 259

```

```

RESULT 13

```


QY 353 C-RGKSDSEYRCYRNQDCEQITSRKGVKMGCTDCEFCAGSYENWIDNRK 407
Db 242 CXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXCLCYACAPYVDWLNQKE 297

RESULT 15
US-08-568-459A-6
Sequence 6, Application US/08568459A
Patent No. 5849306
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chlunis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhuan
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/568,459A
FILING DATE: 07-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelsen, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 749 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
US-08-568-459A-6

Query Match 2.9%; Score 345.5; DB 2; Length 749;
Best Local Similarity 20.3%; Pred. No. 9.2e-15;
Matches 196; Conservative 119; Mismatches 310; Indels 339; Gaps 43;

QY 17 HVLDFPGKIVDEYHG-----EAKIVYSELKGSLSLASTLGTAFTVSKMOTESKYTELIE 72
Db 51 NIFREYKENNVADILFGLNLYEYNNFCKEKPPLVSA-----KY-NLKA 92

QY 73 ANSRNPCKKDGKNDVDFSVKEOAGYDNKKMC--SNGMT-----CAPPRRLHLC-NK 124
Db 93 PNAKSPRIYKKEHESSVFCKTKISKVKKMKWCYSNNKYTKPGVCGPPRROOLCLGY 152

QY 125 NFPNMNSDSSKAKHDLAEYCMAYEGESIKTHYPKYSKYPSPDPMCTMLARSPAD 184
Db 153 IFLRDGNEBELKDH-----INKAANYEAMHLKREY-----ENAGCD-KICNALIGSTAD 201

QY 185 IGDIIIRGDLVIGKKKKKONGKTEREKLEQKLEIFRKIHNDLKDKEAKRYNGDEDPN 244
Db 185 IGDIIIRGDLVIGKKKKKONGKTEREKLEQKLEIFRKIHNDLKDKEAKRYNGDEDPN 244

Db 202 IGDIVRGLDVW-----RDINTNKLSEKFOKIF-----MGGNSRRKQNDNNE-- 243
QY 245 FYKIREDMWTANREYVGMAMTCSKELDSSYFRATQNDTGGPQOTHNKCKCDMDKGNNA 304
Db 244 ----RNKWEKORNLIMSSWV--KHIPKCK--TC-----KRHN----- 274

QY 305 GKPRAGDGVTVTPYFPDYVPOYLRLMFEEMAEDECRRKKKLEMLEKQCRGKDKSDEYRY 364
Db 275 -----FEKIQPLRLKEMGDEFCCEKMGTEVKQLEKICEKNKNSCK-- 315

QY 365 CSRNGYDCEQITSRKGVKMGCGCTDCEFPACSGTSENWIDN-----QRKQKQKATK 417
Db 316 -----KCKNACSSEYKWKERKNEYNLQSKKFFDSKDLINK 350

QY 418 EISDGGKRRKRAVGATTKEGYEKSPYEKLLKNDGYGVADARLGLNNKACKDITDGGKI 477
Db 351 -----KNNLYNKEF-----DSAYLRSEKQCSNIE----- 376

QY 478 NFEKYNSSGGVYGGSGGTSAGSGTNDENKGTG-YRSEYQCPDQGVQHNKGQWERKT 536
Db 377 -----FNDE--TFEPNKKYKACAVC----- 395

QY 537 KYKKRMSKLYKPINGKMWLLLSLKYVKKDMMLKKMKKECLTQNSSDGSVGVTTGA 596
Db 396 --ENPSSSKALKPIKTIVFPIEESK--SELSSLDRSKN--TPNS----- 436

QY 597 SSGNSEKKEL-----YDEMKCYK--HNEVQKYNVQGEV-----EDDELKGA 637
Db 437 GGGNYGDRQIQRDQVHHGDKPEYKSGEYKPIDAAVKTENETSNRNDIEGKESKGD 496

QY 638 GGLCILPNPKKKEVSEAKSONNHADIQTFHDFEYVVAHMLKDSI--HMRTRLKSK 694
Db 497 HSPVHSDIKNEPQRYVSE-NLPKIEKME-----SDSIPYIIEAKGQSS 545

QY 695 ISDGKTAKCRNGCKKCCDCEFKYWKQKETEKPILKDHFKTOEGIPGEYFTTLELLIKLQ 754
Db 546 NSSDNPVAVVSGRSKDVNLHTSERIKENE-----EGV-----IKTDDSSKIE 589

QY 755 FLKEDTEENTENSIDAEAEELKHLQIKLLENENNLAVNAGAEOKTMDKLNHLEND 814
Db 590 ISKIPSDQNNHSDLSQANED-----SNGKNETINPSTENKL--KEIHYKTS 637

QY 815 A-----TKKDCPLPEE-----DK-----SRGSADPSPDIFTPREKEDENE 854
Db 638 SDDHGSKIKSEIEPKELTEESPLTDKTESAIGDKNHESYKSADIF--QSEIHNSDNRD 695

QY 855 DDEDEVARDDEETAKETTESATPT---TTSLDVCPYGVKULTKDNESLDACSLKYG 911
Db 696 RIVSESVVQDSSGSSMSTESIRTDNKDKFTSEDIAPI-----NGR 736

QY 912 NSRL 915
Db 737 NSRV 740

Search completed: May 1, 2001, 13:04:37
Job time: 150 sec

THIS PAGE BLANK (USPTO)